Baskar, P.
10/006780


```
     FILE 'REGISTRY' ENTERED AT 14:27:22 ON 07 OCT 2003
L6          1217 S ATPASE?/CN

     FILE 'HCAPLUS' ENTERED AT 14:27:30 ON 07 OCT 2003
L6          1217 SEA FILE=REGISTRY ABB=ON  PLU=ON  ATPASE?/CN
L7           166 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (PLASMODIUM OR FALCIPARU
                M) AND (L6 OR ATPASE OR ADENOSINE(2W)(TRIPHOSPHATASE OR
                TRI PHOSPHATASE))
L8             4 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L7 AND (MICROTUB? OR
                MICRO TUBUL?)


L6          1217 SEA FILE=REGISTRY ABB=ON  PLU=ON  ATPASE?/CN
L7           166 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (PLASMODIUM OR FALCIPARU
                M) AND (L6 OR ATPASE OR ADENOSINE(2W)(TRIPHOSPHATASE OR
                TRI PHOSPHATASE))
L9            87 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L7 AND (PROTEIN OR
                PEPTIDE OR POLYPROTEIN OR POLYPEPTIDE)
L10            9 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L9 AND STIMUL?


L12           11 L8 OR L10
```

- key terms

L12 ANSWER 1 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2003:454501 HCAPLUS
DOCUMENT NUMBER: 139:35072
TITLE: Vectors comprising nucleotide sequences for target immunogen, PI31, CIIIA and antisense HERNA mRNA, and their uses including use as vaccines
INVENTOR(S): McCreavy, David Thomas; Fraser, William Duncan; Gallagher, James Anthony
PATENT ASSIGNEE(S): University of Liverpool, UK
SOURCE: PCT Int. Appl., 52 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003048371	A2	20030612	WO 2002-GB5512	20021206
WO 2003048371	A3	20030912		

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: GB 2001-29338 A 20011207
GB 2002-23829 A 20021012

AB The invention provides vectors (such as viral vectors, plasmid

THIS PAGE BLANK (USPTO)

vectors or phagemids) comprising: (a) a heterologous nucleotide sequence encoding an antigenic **polypeptide** from a pathogen (such as viral, bacterial, parasitic or fungal); (b) a nucleotide sequence encoding a protease inhibitor (such as human PI31); a nucleotide sequence for a constitutive, regulatable, and/or cell/tissue-specific promoter; and (d) a nucleotide sequence encoding an inhibitory RNA mols., specifically an antisense human HERNA oligonucleotides. The invention also provides vectors comprising a nucleotide sequence encoding CIITA, a **polypeptide** that **stimulates** the expression of MHC class II genes. The invention further provides the use of said vectors as vaccines in production of an immune response (humoral) to said antigens in an animal, such as human, wherein said vaccination may be against a viral, fungal, bacterial or parasitic disorder. Still further, the invention relates: (a) using said vectors in production of antibodies, wherein said antibodies may be of therapeutic and/or of diagnostic use; (b) that said vectors may be adapted for expression of humanized or chimeric antibodies; and (c) that said vectors may be used to used to immunize animals for production of hybridomas expressing a monoclonal antibody against antigen of interest. Finally, the invention provides the cDNA sequences of mouse CIITA, and human PI31, and partial cDNA sequence of human HERNA helicase. The invention related that the use of said vectors containing said sequences can be used to enhance secretion of translated immunogen, and enhance DNA vaccination bias away from an MHC class I event towards MHC class II event. In the examples, the invention presented the construction of two vectors, pcDNAFinal and pcDNA6TR-IRES-CIITA, wherein pcDNAFinal contains nucleotide sequences encoding immunogen parathyroid hormone-related **protein** (PTHrP), antisense HERNA mRNA, PI31 and CD4+ T-cell epitope from lymphocytic choriomeningitis virus and wherein pcDNA6TR-IRES-CIITA encodes CIITA. Specifically, the invention related that: (a) antisense HERNA RNA can increase the transcriptional efficiency of vectors resulting in greater levels of transgene expression; (b) PI31 can inhibit proteasome digestion of recombinant antigen making it more assessable to MHC class II antigens; (c) inclusion of CD4+ T-cell epitope ensured that degraded immunogen-MHC class II complex bound to CD4+ T cells; and (d) inclusion of CIITA **protein** allowed for over-expression of MHC class II antigens.

IT **9000-83-3, Atpase**
RL: BSU (Biological study, unclassified); BIOL (Biological study) (promoter of; vectors comprising sequences for promoter, target immunogen, PI31, CIITA and antisense HERNA mRNA, and their uses, including as vaccines towards various disorders)

L12 ANSWER 2 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2003:454460 HCAPLUS
DOCUMENT NUMBER: 139:31820
TITLE: **Plasmodium falciparum** kinesin motor **protein** KinI-1 with **microtubule-stimulated** **ATPase** activity and uses for diagnosis and treatment of malaria
INVENTOR(S): Sakowicz, Roman; Beraud, Christophe; Guo, Jun; Freedman, Richard
PATENT ASSIGNEE(S): Cytokinetics, Inc., USA
SOURCE: PCT Int. Appl., 68 pp.

CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003048320	A2	20030612	WO 2002-US38360	20021127

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 2003104496	A1	20030605	US 2001-6780	20011130

PRIORITY APPLN. INFO.: US 2001-6780 A2 20011130
US 2002-86935 A2 20020228

AB The present invention relates to **protein** and cDNA sequences of a new kinesin motor **protein**, P. **falciparum** KinI-1 (PfKinI-1), and the use of these compns. for the diagnosis, treatment, or prevention of malaria. The invention provides isolated nucleic acid and amino acid sequences of kinesin superfamily motor **protein** KinI-1, which has **microtubule stimulated ATPase** activity and/or depolymerizes **microtubules**. The invention further relates to antibodies to PfKinI-1, methods of screening for PfKinI-1 modulators using biol. active PfKinI-1, and kits for screening for PfKinI-1 modulators.

IT **9000-83-3, ATPase**
RL: BSU (Biological study, unclassified); BIOL (Biological study) (**microtubule-stimulated; Plasmodium falciparum** kinesin motor **protein** KinI-1 with **microtubule-stimulated ATPase** activity and uses for diagnosis and treatment of malaria)

L12 ANSWER 3 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2003:435206 HCAPLUS
DOCUMENT NUMBER: 139:909
TITLE: Sequences of **Plasmodium falciparum** kinesin KinI-1 and use for treating malaria
INVENTOR(S): Sakowicz, Roman; Beraud, Christophe; Guo, Jun; Freedman, Richard
PATENT ASSIGNEE(S): Cytokinetics, Inc., USA
SOURCE: U.S. Pat. Appl. Publ., 32 pp.
CODEN: USXXCO
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO. KIND DATE APPLICATION NO. DATE

```
---------------  ----  --------      ---------------  --------
US 2003104496     A1   20030605      US 2001-6780     20011130
WO 2003048320     A2   20030612      WO 2002-US38360  20021127
    W:  AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH,
        CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD,
        GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ,
        LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
        NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL,
        TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
        ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
    RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE,
        BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU,
        MC, NL, PT, SE, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
        GW, ML, MR, NE, SN, TD, TG
```

PRIORITY APPLN. INFO.: US 2001-6780 A2 20011130
US 2002-86935 A2 20020228

AB The invention provides sequences of **Plasmodium falciparum** kinesin KinI-1. The invention also relates to the use of KinI-1 for the diagnosis, treatment, or prevention of malaria.

IT **9000-83-3, ATPase**
RL: BSU (Biological study, unclassified); BIOL (Biological study) (**microtubule-stimulated**; sequences of **Plasmodium falciparum** kinesin KinI-1 and use for treating malaria)

L12 ANSWER 4 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2002:752115 HCAPLUS
DOCUMENT NUMBER: 137:289734
TITLE: Sequence of **Plasmodium falciparum** chromosomes 2, 10, 11 and 14
AUTHOR(S): Gardner, Malcolm J.; Shallom, Shamira J.; Carlton, Jane M.; Salzberg, Steven L.; Nene, Vishvanath; Shoaibi, Azadeh; Ciecko, Anne; Lynn, Jeffery; Rizzo, Michael; Weaver, Bruce; Jarrahi, Behnam; Brenner, Michael; Parvizi, Babak; Tallon, Luke; Moazzez, Azita; Granger, David; Fujii, Claire; Hansen, Cheryl; Pederson, James; Feldblyum, Tamara; Peterson, Jeremy; Suh, Bernard; Angiuoli, Sam; Pertea, Mihaela; Allen, Jonathan; Selengut, Jeremy; White, Owen; Cummings, Leda M.; Smith, Hamilton O.; Adams, Mark D.; Venter, J. Craig; Carucci, Daniel J.; Hoffman, Stephen L.; Fraser, Claire M.
CORPORATE SOURCE: The Institute for Genomic Research, Rockville, MD, 20850, USA
SOURCE: Nature (London, United Kingdom) (2002), 419(6906), 531-534
CODEN: NATUAS; ISSN: 0028-0836
PUBLISHER: Nature Publishing Group
DOCUMENT TYPE: Journal
LANGUAGE: English

AB The mosquito-borne malaria parasite **Plasmodium falciparum** kills an estimated 0.7-2.7 million people every year, primarily children in sub-Saharan Africa. Without effective interventions, a variety of factors-including the spread of parasites resistant to antimalarial drugs and the increasing insecticide resistance of mosquitoes-may cause the number of malaria

cases to double over the next two decades. To **stimulate** basic research and facilitate the development of new drugs and vaccines, the genome of **Plasmodium falciparum** clone 3D7 has been sequenced using a chromosome-by-chromosome shotgun strategy. This report describes nucleotide sequences of chromosomes 10, 11 and 14, and a re-anal. of the chromosome 2 sequence. These chromosomes represent about 35% of the 23-megabase P. **falciparum** genome. The sequences are deposited in GenBank/EMBL/DDBJ under accession nos. AE001362.2 (chromosome 2), AE014185 (chromosome 10), AE014186 (chromosome 11), and AE014187 (chromosome 14).

only NA < no other

IT **465605-54-3**
RL: BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)
(amino acid sequence; complete sequence of **Plasmodium falciparum** chromosomes 2, 10, 11 and 14)

REFERENCE COUNT: 30 THERE ARE 30 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 5 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER:	2002:485238 HCAPLUS
DOCUMENT NUMBER:	137:242895
TITLE:	Comparative genomic analysis in the region of a major **Plasmodium**-refractoriness locus of Anopheles gambiae
AUTHOR(S):	Thomasova, Dana; Ton, Lucas Q.; Copley, Richard R.; Zdobnov, Evgeny M.; Wang, Xuelan; Hong, Young S.; Sim, Cheolho; Bork, Peer; Kafatos, Fotis C.; Collins, Frank H.
CORPORATE SOURCE:	European Molecular Biology Laboratory, Heidelberg, 69117, Germany
SOURCE:	Proceedings of the National Academy of Sciences of the United States of America (2002), 99(12), 8179-8184 CODEN: PNASA6; ISSN: 0027-8424
PUBLISHER:	National Academy of Sciences
DOCUMENT TYPE:	Journal
LANGUAGE:	English

AB We have sequenced six overlapping clones from a library of bacterial artificial chromosome (BAC) clones derived from a laboratory strain of the mosquito, Anopheles gambiae, the major vector of human malaria in Africa. The resulting uninterrupted 528-kb sequence is from the 8C region of the mosquito 2R chromosome, at or very near the major refractoriness locus associated with melanotic encapsulation of parasites. This sequence represents the first extensive view of the mosquito genome structure encompassing 48 genes. Genomic comparison reveals that the majority of the orthologues are found in six microsyntenic clusters in Drosophila melanogaster. A BAC clone that is wholly contained within this region demonstrates the existence of a remarkable degree of local polymorphism in this species, which may prove important for its population structure and vectorial capacity.

IT **9000-83-3**
RL: BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)
(V-type, sequence homolog to; comparative genomic anal. in the region of a major **Plasmodium**-refractoriness locus of Anopheles gambiae)

REFERENCE COUNT: 40 THERE ARE 40 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 6 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2001:394971 HCAPLUS
DOCUMENT NUMBER: 135:118715
TITLE: Expression and functional characterization of a **Plasmodium falciparum** Ca2+-**ATPase** (PfATP4) belonging to a subclass unique to apicomplexan organisms
AUTHOR(S): Krishna, Sanjeev; Woodrow, Charles; Webb, Richard; Penny, Jeff; Takeyasu, Kunio; Kimura, Masatsugu; East, J. Malcolm
CORPORATE SOURCE: Department of Infectious Diseases, St. George's Hospital Medical School, London, SW17 ORE, UK
SOURCE: Journal of Biological Chemistry (2001), 276(14), 10782-10787
CODEN: JBCHA3; ISSN: 0021-9258
PUBLISHER: American Society for Biochemistry and Molecular Biology
DOCUMENT TYPE: Journal
LANGUAGE: English
AB We have obtained a full-length P type **ATPase** sequence (PfATP4) encoded by **Plasmodium falciparum** and expressed PfATP4 in Xenopus laevis oocytes to study its function. Comparison of the hitherto incomplete open reading frame with other Ca2+-**ATPase** sequences reveals that PfATP4 differs significantly from previously defined categories. The Ca2+-dependent **ATPase** activity of PfATP4 is **stimulated** by a much broader range of [Ca2+]free (3.2-320 μM) than are an avian SERCA1 pump or rabbit SERCA 1a (maximal activity < 10 μM). The activity of PfATP4 is resistant to inhibition by ouabain (200 μM) or thapsigargin (0.8 μM) but is inhibited by vanadate (1 mM) or cyclopiazonic acid (1 μM). We used a quant. polymerase chain reaction to assay expression of mRNA encoding PfATP4 relative to that for β-tubulin in synchronized asexual stages and found variable expression throughout the life cycle with a maximal 5-fold increase in meronts compared with ring stages. This anal. suggests that PfATP4 defines a novel subclass of Ca2+-**ATPases** unique to apicomplexan organisms and therefore offers potential as a drug target.
IT **9000-83-3, ATPase**
RL: BAC (Biological activity or effector, except adverse); BPR (Biological process); BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study); PROC (Process)
(calcium-activated, PfATP4; expression and functional characterization of a **Plasmodium falciparum** Ca2+-**ATPase** (PfATP4) belonging to a subclass unique to apicomplexan organisms)
REFERENCE COUNT: 38 THERE ARE 38 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L12 ANSWER 7 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1994:694055 HCAPLUS
DOCUMENT NUMBER: 121:294055
TITLE: **Plasmodium falciparum:**

further characterization of putative cation **ATPases**
AUTHOR(S): Krishna, Sanjeev; Cowan, Gill M.; Robson, Kathryn J.; Meade, John C.
CORPORATE SOURCE: Inst. Mol. Med., John Radcliffe Hosp., Oxford, OX3 9DU, UK
SOURCE: Experimental Parasitology (1994), 78(1), 113-17
CODEN: EXPAAA; ISSN: 0014-4894
DOCUMENT TYPE: Journal
LANGUAGE: English

AB The emergence of multi-drug resistant strains of **Plasmodium falciparum** and the current lack of a vaccine have **stimulated** research to identify new chemotherapeutic targets. The authors have focused on the isolation and characterization of nucleotide sequences encoding putative "P"-type cation **ATPases** from P. **falciparum**, because selective inhibitors of mammalian members of this transporter family are well studied, and some are already in clin. use. This report describes the isolation and sequence anal. of two addnl. members of this family, **ATPases** 2 and 3, and their differential expressivity during the erythrocytic stage of parasite development. Amino acid sequence anal. identified the 7 amino acid motifs present at the phosphorylation site (DKTGTLT) of other P-type cation **ATPases**, in both clones. **ATPase** 2 was localized to chromosome 5 and **ATPase** 3 to chromosome 12 on a Southern blot of chromosomes resolved by pulsed-field gel electrophoresis. Northern blot anal. showed that both **ATPases** 2 and 3 were expressed during the erythrocytic stage of the infection. **ATPase** 2 has a mRNA transcript size of 5.5-6 kb and is expressed at the schizont stage of development. **ATPase** 3 mRNA is approx. 5 kb in size and the strongest signal corresponds to mRNA from parasites at the ring stage of development (12-24 h postinvasion), although mRNA is also detected in all other erythrocytic stages examined

IT **9000-83-3, ATPase**
RL: PRP (Properties)
(2 and 3; sequence and chromosomal localization and developmental expression of genes encoding P-type cation **ATPases** 2 and 3 of **Plasmodium falciparum**)

L12 ANSWER 8 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1994:407323 HCAPLUS
DOCUMENT NUMBER: 121:7323
TITLE: Immunomodulatory **peptides** binding to human major histocompatibility complex (MHC) class II allotype
INVENTOR(S): Urban, Robert Glen; Chicz, Roman M.; Vignali, Dario A. A.; Hedley, Mary Lynne; Stern, Lawrence J.; Strominger, Jack L.
PATENT ASSIGNEE(S): President and Fellows of Harvard College, USA
SOURCE: PCT Int. Appl., 59 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE

```
     ---------------  ----  --------      ---------------  --------
     WO 9404557       A1    19940303      WO 1992-US6692   19920811
         W:  JP
     JP 08502244      T2    19960312      JP 1992-506181   19920811
PRIORITY APPLN. INFO.:                  WO 1992-US6692     19920811
```

AB   A purified oligopeptide preparation comprising an amino acid sequence identical to that of a segment of a naturally-occurring human **protein** that binds to human major histocompatibility complex (MHC) class II allotype is provided. The human **protein** is an MHC class I or II mol., HLA-A2, invariant chain (Ii), etc.,. A method is described for inhibiting an immune response in a human patient by contacting an antigen-presenting cell (APC) of the patient with a therapeutic composition or an immune-**stimulating** complex (ISCOM) containing the oligopeptide, or by expression of the oligopeptide-coding sequence linked to a trafficking sequence in APCs. The oligopeptide also can be used for inducing an immune response against pathogens. The options of the oligopeptide delivery system is also described. Purification and characterization of 6 HLA-DR antigens (HLA-DR1.apprx.4; HLA-DR7.apprx.8) from Epstein-Barr virus-transformed human B lymphoblastoid cell lines were demonstrated.

IT   **9000-83-3, ATPase**
     RL: BIOL (Biological study)
        (Na+/K+, **peptides** of, as immunomodulators)

L12 ANSWER 9 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1991:673744 HCAPLUS
DOCUMENT NUMBER: 115:273744
TITLE: **Stimulation** of the interaction between actin and myosin by Physarum caldesmon-like **protein** and smooth muscle caldesmon
AUTHOR(S): Ishikawa, Ryoki; Okagaki, Tsuyoshi; Higashi-Fujime, Sugie; Kohama, Kazuhiro
CORPORATE SOURCE: Sch. Med., Gunma Univ., Maebashi, 371, Japan
SOURCE: Journal of Biological Chemistry (1991), 266(32), 21784-90
CODEN: JBCHA3; ISSN: 0021-9258
DOCUMENT TYPE: Journal
LANGUAGE: English

AB   An actin-binding **protein** was purified from the **plasmodia** of a lower eukaryote, P. polycephalum, with an apparent mol. weight of 210 kDa on SDS-PAGE. This **protein** bound to actin filaments with a stoichiometry of 1:7-8 in a Ca2+-calmodulin-dependent manner. Antibody raised against caldesmon from smooth muscle cross-reacted with the 210-kDa **protein.** In vitro motility assay revealed that the 210-kDa **protein** increased the sliding velocity of actin filaments on Physarum myosin. The 210-kDa **protein** more than doubled the actin-activated **ATPase** activity of Physarum myosin under comparative conditions of in vitro motility assay. Further increases in the concentration of the 210-kDa **protein** decreased its **stimulatory** effects. Ca2+-calmodulin prevented the **stimulatory** effects of the 210-kDa **protein.** Unexpectedly, smooth muscle caldesmon also increased the sliding velocity of actin filaments on smooth muscle myosin at lower concns. The well-known inhibitory effect of smooth muscle caldesmon on the actin-myosin interaction was observed with this motility assay when the concentration of the caldesmon was increased further. The

stimulatory and inhibitory effects were confirmed by measurements of actin-activated **ATPase** activity of smooth muscle myosin. From estns. of the intracellular concns. of the 210-kDa **protein** and smooth muscle caldesmon in vivo, it appears that effects of the former and the latter on actin-myosin interactions in vivo are **stimulatory** and inhibitory, resp.

IT **9000-83-3, ATPase**
RL: BIOL (Biological study)
(of myosin, actin-binding **protein** of Physarum polycephalum and caldesmon of smooth muscle effect on)

L12 ANSWER 10 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1989:3408 HCAPLUS
DOCUMENT NUMBER: 110:3408
TITLE: Purification of **Plasmodium falciparum** digestive vacuoles and partial characterization of the vacuolar membrane **ATPase**
AUTHOR(S): Choi, Inpyo; Mego, John L.
CORPORATE SOURCE: Dep. Biol., Univ. Alabama, Tuscaloosa, AL, USA
SOURCE: Molecular and Biochemical Parasitology (1988), 31(1), 71-8
CODEN: MBIPDP; ISSN: 0166-6851
DOCUMENT TYPE: Journal
LANGUAGE: English

AB P. **falciparum** digestive vacuoles containing Fe^{3+} oxide granules were purified from parasite homogenates by centrifugation on discontinuous sucrose gradients. Digestive vacuole membranes prepared by osmotic lysis and washed with KCl showed no detectable contamination by erythrocyte membrane **proteins** and only minimal contamination by nonvacuolar parasite **proteins**. Purified vacuolar membranes were 2.6-fold enriched in total parasite membrane **ATPase** activity. This **ATPase** was optimally active at pH 7 in the presence of >2 mM Mg^{2+}. Ca^{2+} and Mn^{2+} were .apprx.80-90% as effective as Mg^{2+}, and Zn^{2+}, Co^{2+}, and Fe^{2+} also exerted some **stimulatory** effect. The vacuolar membrane also hydrolyzed GTP, UTP, CTP, and ADP, but AMP and 3',5'-cAMP were hydrolyzed only one-tenth as effectively as ATP. The **ATPase** was unaffected by vanadate, ouabain, or oligomycin but was significantly inhibited by the proton pump inhibitors NEM and NBD-Cl. Of 6 antimalarial drugs tested, quinine and quinacrine were the most effective inhibitors and mefloquine was the least effective.

IT **9000-83-3P, ATPase**
RL: PREP (Preparation)
(of digestive vacuoles of **Plasmodium falciparum**, purification and characterization of)

L12 ANSWER 11 OF 11 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 1982:434955 HCAPLUS
DOCUMENT NUMBER: 97:34955
TITLE: A novel 36,000-dalton actin-binding protein purified from microfilaments in Physarum **plasmodia** which aggregates actin filaments and blocks actin-myosin interaction
AUTHOR(S): Ogihara, Satoshi; Tonomura, Yuji
CORPORATE SOURCE: Fac. Sci., Osaka Univ., Osaka, 560, Japan
SOURCE: Journal of Cell Biology (1982), 93(3), 604-14

CODEN: JCLBA3; ISSN: 0021-9525
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The **plasmodia** of P. polycephalum contains large aggregates of entangled actin microfilaments. Treatment with Triton X-100 gave a demembraned cytoskeleton consisting of entangled actin filaments which showed almost no interaction with rabbit skeletal myosin. A novel actin-binding protein purified from the cytoskeleton stoichiometrically binds to actin and causes actin filaments to curl and aggregate. The protein inhibits **ATPase** activity as well as the superpptn. of reconstituted rabbit skeletal muscle actomyosin. This protein has a mol. weight of 36,000 and binds 7 mol of actin/mol 36,000 polypeptide.
IT **9000-83-3**
RL: BIOL (Biological study)
(inhibition of, of actomyosin, actin-binding protein inhibition of)

(FILE 'MEDLINE, BIOSIS, EMBASE, WPIDS, CONFSCI, SCISEARCH, JICST-EPLUS, JAPIO' ENTERED AT 14:31:47 ON 07 OCT 2003)

```
L13            3 S L8
L14           21 S L10
L15           24 S L13 OR L14
L16           17 DUP REM L15 (7 DUPLICATES REMOVED)
```

L16 ANSWER 1 OF 17 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2003-505298 [47] WPIDS
DOC. NO. CPI: C2003-135135
TITLE: New vector, useful for preparing a composition for treating or preventing bacterial, viral, fungal or parasitic infection.
DERWENT CLASS: B04 C06 D16
INVENTOR(S): FRASER, W D; GALLAGHER, J A; MCCREAVY, D T
PATENT ASSIGNEE(S): (UYLI-N) UNIV LIVERPOOL
COUNTRY COUNT: 101
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2003048371	A2	20030612	(200347)*	EN	52

RW: AT BE BG CH CY CZ DE DK EA EE ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SI SK SL SZ TR TZ UG ZM ZW
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SC SD SE SG SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM ZW

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2003048371	A2	WO 2002-GB5512	20021206

PRIORITY APPLN. INFO: GB 2002-23829 20021012; GB 2001-29338 20011207
AN 2003-505298 [47] WPIDS

AB WO2003048371 A UPAB: 20030723
NOVELTY - An vector comprising a heterologous nucleic acid sequence encoding an antigenic **polypeptide** and a nucleic acid molecule comprising a 3188 base pair sequence, given in the specification, a nucleic acid molecule which hybridizes to it and which encodes a protease inhibitor **polypeptide**, or nucleic acid molecules which comprise degenerate nucleic acid sequences. The vector is adapted for the expression of each **polypeptide**.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) inducing an immune response to an antigenic **polypeptide**;
(2) an antibody;
(3) a cell transformed with the novel vector;
(4) producing humanized or chimeric antibody;
(5) a hybridoma cell line which produces a monoclonal antibody;
(6) a vaccine comprising the novel vector; and
(7) vaccinating an animal, preferably a human, against at least one pathological condition.

ACTIVITY - Antibacterial; Virucide; Antiparasitic; Antifungal; Anti-HIV; Antiulcer.

No biological data is given.

MECHANISM OF ACTION - Gene therapy; Vaccine.

USE - The vector is useful for preparing a composition for preventing or treating AIDS, herpes, rubeola, rubella, varicella, influenza, common cold or viral meningitis; septicemia, tuberculosis, bacteria-associated food poisoning, blood infections, peritonitis, endocarditis, sepsis, bacterial meningitis, pneumonia, stomach ulcers, gonorrhea, strep throat, streptococcal-associated toxic shock, necrotizing fasciitis, impetigo, histoplasmosis, Lyme disease, gastro-enteritis, dysentery or shigellosis; Candidiasis; or trypanosomiasis, malaria, schistosomiasis or Chagas disease (claimed).
Dwg.0/9

L16 ANSWER 2 OF 17 EMBASE on STN

ACCESSION NUMBER:	2003300235 EMBASE
TITLE:	The pathophysiology of **falciparum** malaria.
AUTHOR:	Clark I.A.; Cowden W.B.
CORPORATE SOURCE:	I.A. Clark, Sch. of Biochem./Molecular Biology, Australian National University, Canberra, ACT 0200, Australia. ian.clark@anu.edu.au
SOURCE:	Pharmacology and Therapeutics, (1 Aug 2003) 99/2 (221-260). Refs: 466 ISSN: 0163-7258 CODEN: PHTHDT
COUNTRY:	United States
DOCUMENT TYPE:	Journal; General Review
FILE SEGMENT:	004 Microbiology
	017 Public Health, Social Medicine and Epidemiology
	026 Immunology, Serology and Transplantation
	037 Drug Literature Index
	038 Adverse Reactions Titles
LANGUAGE:	English
SUMMARY LANGUAGE:	English

AB **Falciparum** malaria is a complex disease with no simple

explanation, affecting organs where the parasite is rare as well as those organs where it is more common. We continue to argue that it can best be understood in terms of excessive **stimulation** of normally useful pathways mediated by inflammatory cytokines, the prototype being tumor necrosis factor (TNF). These pathways involve downstream mediators, such as nitric oxide (NO) that the host normally uses to control parasites, but which, when uncontrolled, have bioenergetic failure of patient tissues as their predictable end point. **Falciparum** malaria is no different from many other infectious diseases that are clinically confused with it. The sequestration of parasitized red blood cells, prominent in some tissues but absent in others with equal functional loss, exacerbates, but does not change, these overriding principles. Recent opportunities to stain a wide range of tissues from African pediatric cases of **falciparum** malaria and sepsis for the inducible NO synthase (iNOS) and migration inhibitory factor (MIF) have strengthened these arguments considerably. The recent demonstration of bioenergetic failure in tissue removed from sepsis patients being able to predict a fatal outcome fulfils a prediction of these principles, and it is plausible that this will be demonstrable in severe **falciparum** malaria. Understanding the disease caused by **falciparum** malaria at a molecular level requires an appreciation of the universality of poly(ADP-ribose) polymerase-1 (PARP-1) and Na(+)/K(+)-**ATPase** and the protean effects of activation by inflammation of the former that include inactivation of the latter.

L16 ANSWER 3 OF 17 EMBASE on STN DUPLICATE 1

ACCESSION NUMBER: 2003191572 EMBASE
TITLE: Replication fork-**stimulated** eIF-4A from **Plasmodium** cynomolgi unwinds DNA in the 3' to 5' direction and is inhibited by DNA-interacting compounds.
AUTHOR: Tuteja R.; Tuteja N.; Malhotra P.; Chauhan V.S.
CORPORATE SOURCE: R. Tuteja, Intl. Ctr. for Genetic Eng./Biotech., Aruna Asaf Ali Marg, New Delhi 110067, India. renu@icgeb.res.in
SOURCE: Archives of Biochemistry and Biophysics, (1 Jun 2003) 414/1 (108-114).
Refs: 47
ISSN: 0003-9861 CODEN: ABBIA4
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 030 Pharmacology
037 Drug Literature Index
LANGUAGE: English
SUMMARY LANGUAGE: English

AB **Plasmodium** cynomolgi DEAD-box DNA helicase 45 (PcDDH45) is an ATP-dependent DNA-unwinding enzyme with intrinsic DNA-dependent **ATPase** activity and is highly homologous to eIF-4A. In this study, we have further characterized and tested the effect of various DNA-interacting compounds on the DNA-unwinding activity of PcDDH45. The results show that PcDDH45 translocates in the 3' to 5' direction along the bound strand, a replication fork-like structure of the substrate **stimulates** its DNA-unwinding activity, and it failed to unwind blunt-ended duplex DNA. Of various compounds

tested, only cisplatin, 4',6'-diamidino-2-phenylindole, daunorubicin, and nogalamycin were inhibitory to the unwinding activity of PcDDH45 with apparent IC(50) values of 1.0, 4.0, 7.5, and 1.7μM, respectively. These results suggest that the interaction of these compounds with duplex DNA generate a complex that probably impedes the translocation of PcDDH45, resulting in inhibition of unwinding activity. This study is one of the first to demonstrate the effect of various DNA-binding compounds on a malaria parasite DNA helicase and should make an important contribution to our better understanding of the nucleic acid transactions in the parasite.

L16 ANSWER 4 OF 17 SCISEARCH

ACCESSION NUMBER: 2002:547065 SCISEARCH
THE GENUINE ARTICLE: 566XK
TITLE: Multidrug resistance phenotype mediated by the P-glycoprotein-like transporter in Leishmania: A search for reversal agents
AUTHOR: Perez-Victoria J M; Di Pietro A; Barron D; Ravelo A G; Castanys S; Gamarro F (Reprint)
CORPORATE SOURCE: CSIC, Inst Parasitol & Biomed Lopez Neyra, C Ventanilla 11, Granada 18001, Spain (Reprint); CSIC, Inst Parasitol & Biomed Lopez Neyra, Granada 18001, Spain; CNRS, UMR 5086, Inst Biol & Chim Prot, Lyon, France; Univ Lyon 1, CNRS, Lab Prod Nat, F-69622 Villeurbanne, France; Univ La Laguna, Inst Bioorgan Antonio Gonzalez, Tenerife, Spain
COUNTRY OF AUTHOR: Spain; France
SOURCE: CURRENT DRUG TARGETS, (AUG 2002) Vol. 3, No. 4, pp. 311-333.
Publisher: BENTHAM SCIENCE PUBL LTD, PO BOX 1673, 1200 BR HILVERSUM, NETHERLANDS.
ISSN: 1389-4501.
DOCUMENT TYPE: General Review; Journal
LANGUAGE: English
REFERENCE COUNT: 193
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Protozoan parasites are responsible for important diseases that threaten the lives of nearly one-quarter of the human population world-wide. Among them, leishmaniasis has become the second cause of death, mainly due to the emergence of parasite resistance to conventional drugs. P-glycoprotein (Pgp)-like transporters overexpression is a very efficient mechanism to reduce the intracellular accumulation of many drugs in cancer cells and parasitic protozoans including **Plasmodium** and Leishmania, thus conferring a multidrug resistance (MDR) phenotype. Therefore, there is a great clinical interest in developing inhibitors of these transporters to overcome such a resistance. Pgps are active pumps belonging to the ATP-binding cassette (ABC) superfamily of **proteins**, and consist of two homologous halves, each containing a transmembrane domain (TMD) involved in drug efflux, and a cytosolic nucleotide-binding domain (NBD) responsible for ATP binding and hydrolysis. Most conventional cancer MDR modulators interact with the drug-binding sites on the TMDs of Pgps, but they are also usually transported and the required concentrations for a permanent inhibition produce subsequent side-effects that hamper their clinical use. Besides, they only poorly modulate the

resistance in protozoan parasites. We review here a rational strategy developed to overcome the MDR phenotype in Leishmania, consisting in: i) the selection of an MDR Leishmania tropica line that overexpresses a Pgp-like transporter; ii) the use of their cytosolic NBDs as new pharmacological targets; iii) the search of new natural compounds that revert the MDR phenotype in Leishmania by binding to the TMDs; iv) the combination of subdoses of the above selected modulators directed to both targets in the transporter, NBDs and TMDs, to accumulate their reversal effects while diminishing their toxicity. In this way, we have reverted the MDR phenotype in Leishmania, including the resistance to the most promising new antileishmania agents, the alkyl-lysophospholipids. This approach might be extrapolated to be used in other eukaryotic cells.

L16 ANSWER 5 OF 17 EMBASE on STN
ACCESSION NUMBER: 2002097347 EMBASE
TITLE: Cytoskeleton of apicomplexan parasites.
AUTHOR: Morrissette N.S.; Sibley L.D.
CORPORATE SOURCE: N.S. Morrissette, Department of Molecular Microbiology, Washington Univ. School of Medicine, 660 South Euclid Ave., St. Louis, MO 63110, United States. naomi@borcim.wustl.edu
SOURCE: Microbiology and Molecular Biology Reviews, (2002) 66/1 (21-38).
Refs: 200
ISSN: 1092-2172 CODEN: MMBRF7
COUNTRY: United States
DOCUMENT TYPE: Journal; General Review
FILE SEGMENT: 004 Microbiology
030 Pharmacology
037 Drug Literature Index
LANGUAGE: English
SUMMARY LANGUAGE: English
AB The Apicomplexa are a phylum of diverse obligate intracellular parasites including **Plasmodium** spp., the cause of malaria; Toxoplasma gondii and Cryptosporidium parvum, opportunistic pathogens of immunocompromised individuals; and Eimeria spp. and Theileria spp., parasites of considerable agricultural importance. These protozoan parasites share distinctive morphological features, cytoskeletal organization, and modes of replication, motility, and invasion. This review summarizes our current understanding of the cytoskeletal elements, the properties of cytoskeletal proteins, and the role of the cytoskeleton in polarity, motility, invasion, and replication. We discuss the unusual properties of actin and myosin in the Apicomplexa, the highly stereotyped **microtubule** populations in apicomplexans, and a network of recently discovered novel intermediate filament-like elements in these parasites.

L16 ANSWER 6 OF 17 MEDLINE on STN DUPLICATE 2
ACCESSION NUMBER: 2001560947 MEDLINE
DOCUMENT NUMBER: 21519010 PubMed ID: 11606229
TITLE: **Microtubule** associated motor proteins of **Plasmodium falciparum** merozoites.
AUTHOR: Fowler R E; Smith A M; Whitehorn J; Williams I T; Bannister L H; Mitchell G H
CORPORATE SOURCE: Malaria Laboratory, Department of Immunobiology,

Guy's, King's and St Thomas' School of Medicine, KCL,
Guy's Hospital, London Bridge, London, SE1 9RT, UK..
ruth.fowler@ed.ac.uk

SOURCE: MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (2001 Oct) 117 (2) 187-200.
Journal code: 8006324. ISSN: 0166-6851.
PUB. COUNTRY: Netherlands
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200201
ENTRY DATE: Entered STN: 20011022
Last Updated on STN: 20020201
Entered Medline: 20020131

AB We have studied the occurrence, stage specificity and cellular location of key molecules associated with **microtubules** in **Plasmodium falciparum** merozoites. Antibodies to gamma tubulin, conventional kinesin and cytoplasmic dynein were used to determine the polarity of merozoite **microtubules** (mt), the stage specificity of the motor proteins and their location during merozoite development. We conclude that the minus ends of the mts are located at their apical pole. Kinesin was present throughout the lifecycle, appearing as a distinct crescent at the apex of developing merozoites. The vast majority of cytoplasmic dynein reactivity occurred in late merogony, also appearing at the merozoite apex. Destruction of mt with dinitroanilines did not affect the cellular location of kinesin or dynein. In invasion assays, dynein inhibitors reduced the number of ring stage parasites. Our results show that both conventional kinesin and cytoplasmic dynein are abundant, located at the negative pole of the merozoite mt and, intriguingly, appear there only in very late merogony, prior to merozoite release and invasion.

L16 ANSWER 7 OF 17 MEDLINE on STN

ACCESSION NUMBER: 2001297917 MEDLINE
DOCUMENT NUMBER: 21273137 PubMed ID: 11378198

TITLE: Two classes of plant-like vacuolar-type H(+)-pyrophosphatases in malaria parasites.

AUTHOR: McIntosh M T; Drozdowicz Y M; Laroiya K; Rea P A; Vaidya A B
CORPORATE SOURCE: Department of Microbiology and Immunology, MCP Hahnemann University, 2900 Queen Lane, Philadelphia, PA 19129, USA.. michael.t.mcintosh@drexel.edu

CONTRACT NUMBER: AF28398
SOURCE: MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (2001 May) 114 (2) 183-95.
Journal code: 8006324. ISSN: 0166-6851.
PUB. COUNTRY: Netherlands
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
OTHER SOURCE: GENBANK-AF115766; GENBANK-AF115767; GENBANK-AF283528
ENTRY MONTH: 200108
ENTRY DATE: Entered STN: 20010806
Last Updated on STN: 20010806
Entered Medline: 20010802

AB In plants, cytosolic inorganic pyrophosphate (PP(i)) is hydrolyzed by energy-conserving vacuolar-type H(+)-pyrophosphatases (V-PPases)

that harness the free energy of PP(i) hydrolysis to establish transmembrane H(+) gradients. Here we describe the identification and cloning of two genes, PfVP1 and PfVP2, from the malaria parasite **Plasmodium falciparum**. Inferred to encode type I (K(+)-dependent) and type II (K(+)-independent) V-PPases, respectively, PfVP1 and PfVP2 appeared more sequence divergent from each other than from their type I and type II counterparts in plants. The steady state levels of PfVP1 mRNA were high in comparison to PfVP2 mRNA throughout the erythrocytic phases of infection. Western analyses of trophozoite membranes using generic V-PPase antibodies (PAB(HK) and PAB(TK)) demonstrated appreciable amounts of a Mr 67000 **polypeptide** whose associated aminomethylenediphosphonate- (AMDP) inhibitable PPase activity was markedly **stimulated** by K(+). Immunofluorescence microscopy of infected erythrocytes revealed PfVP antigen associated with both the parasite plasma membrane and punctate intracellular inclusions. Transient transfection of a PfVP1-GFP fusion further supported the localization of PfVP1 to the parasite plasma membrane. Based on these findings and the growth-retarding effects of AMDP, P. **falciparum** is concluded to possess both type I and type II V-PPases of which the former has the greatest potential for contributing to the establishment of H(+) gradients across the parasite plasma membrane under conditions of energy limitation.

L16 ANSWER 8 OF 17 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2001:373496 SCISEARCH
THE GENUINE ARTICLE: 427QR
TITLE: Detection and localization of a Ca2+-**ATPase** activity in Toxoplasma gondii
AUTHOR: Bouchot A; Jaillet J D; Bonhomme A (Reprint); Pezzella-D'Alessandro N; Laquerriere P; Kilian L; Burlet H; Gomez-Marin J E; Pluot M; Bonhomme P; Pinon J M
CORPORATE SOURCE: CHU Maison Blanche, IFR 53, UPRES EA 2070, 51 Rue Cognacq Jay, F-51095 Reims, France (Reprint); CHU Maison Blanche, IFR 53, UPRES EA 2070, F-51095 Reims, France; CHU Maison Blanche, IFR 53, Lab Parasitol Mycol, F-51095 Reims, France; Lab Microscopie Elect Analyt & Quantitat, IFR 53, Reims, France; Hop Robert Debre, Anat Pathol Lab, Reims, France; Univ Nacl Colombia, Hosp San Juan de Dios, Dept Med Interna, Grp Pathol Infecciosa, Santafe De Bogota, Colombia
COUNTRY OF AUTHOR: France; Colombia
SOURCE: CELL STRUCTURE AND FUNCTION, (FEB 2001) Vol. 26, No. 1, pp. 49-60.
Publisher: JAPAN SOC CELL BIOLOGY, SHIMOTACHIURI OGAWA-HIGASHI, KAMIKYOKU KYOTO, 602, JAPAN.
ISSN: 0386-7196.
DOCUMENT TYPE: Article; Journal
LANGUAGE: English
REFERENCE COUNT: 59
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Toxoplasma gondii, the agent causing toxoplasmosis, is an obligate intracellular protozoan parasite. A calcium signal appears to be essential for intracellular transduction during the active process of host cell invasion. We have looked for a Ca2+-transport **ATPase** in tachyzoites and found Ca2+-**ATPase**

activity (11-22 nmol Pi liberated/mg **protein**/min) in the tachyzoite membrane fraction, This ATP-dependent activity was **stimulated** by Ca2+ and Mg2+ ions and by calmodulin, and was inhibited by pomp inhibitors (sodium orthovanadate or thapsigargin). We used cytochemistry and X-ray microanalysis of cerium phosphate precipitates and immunolabelling to find the Ca2+, Mg2+-**ATPase**. It was located mainly in the membrane complex, the conoid, nucleus, secretory organelles (rhoptries, dense granules) and in vesicles with a high calcium concentration, Thus, Toxoplasma gondii possesses Ca2+-pump **ATPase** (Ca2+, Mg2+-**ATPase**) as do eukaryotic cells.

L16 ANSWER 9 OF 17 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN

ACCESSION NUMBER:	2000:622594 SCISEARCH
THE GENUINE ARTICLE:	343QA
TITLE:	Vacuolar proton pyrophosphatase activity and pyrophosphate (PPi) in Toxoplasma gondii as possible chemotherapeutic targets
AUTHOR:	Rodrigues C O; Scott D A; Bailey B N; deSouza W; Benchimol M; Moreno B; Urbina J A; Oldfield E; Moreno S N J (Reprint)
CORPORATE SOURCE:	UNIV ILLINOIS, COLL VET MED, DEPT PATHOBIOL, MOL PARASITOL LAB, 2001 S LINCOLN AVE, URBANA, IL 61802 (Reprint); UNIV ILLINOIS, COLL VET MED, DEPT PATHOBIOL, MOL PARASITOL LAB, URBANA, IL 61802; UNIV ILLINOIS, DEPT CHEM, URBANA, IL 61801; UNIV ILLINOIS, DEPT BIOPHYS, URBANA, IL 61801; UNIV SANTA URSULA, RIO JANEIRO, BRAZIL; INST VENEZOLANO INVEST CIENT, CARACAS, VENEZUELA; UNIV FED RIO DE JANEIRO, INST BIOFIS CARLOS CHAGAS FILHO, BR-21941 RIO JANEIRO, BRAZIL
COUNTRY OF AUTHOR:	USA; BRAZIL; VENEZUELA
SOURCE:	BIOCHEMICAL JOURNAL, (1 AUG 2000) Vol. 349, Part 3, pp. 737-745. Publisher: PORTLAND PRESS, 59 PORTLAND PLACE, LONDON W1N 3AJ, ENGLAND. ISSN: 0264-6021.
DOCUMENT TYPE:	Article; Journal
FILE SEGMENT:	LIFE
LANGUAGE:	English
REFERENCE COUNT:	50

ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The addition of PPi promoted the acidification of a subcellular compartment in cell homogenates of Toxoplasma gondii tachyzoites, implying the presence of a proton-translocating pyrophosphatase. The proton gradient was collapsed by addition of the K+/H+ antiporter nigericin, and was also inhibited by addition of the PP, analogue aminomethylenediphosphonate (AMDP). Both proton transport and PPi hydrolysis were dependent upon K+, but Na+ caused partial inhibition of these activities. PPi hydrolysis was sensitive in a dose-dependent manner to AMDP, imidodiphosphate, NaF and to the thiol reagent N-ethylmaleimide. This activity was unaffected by common inhibitors of phosphohydrolases, except that NaO3V (sodium orthovanadate) **stimulated** the activity by 87%. Immunofluorescence microscopy, using antisera raised against conserved **peptide** sequences of a plant vacuolar pyrophosphatase, suggested that the pyrophosphatase in T. gondii tachyzoites was located in the plasma membrane and intracellular

vacuoles of the parasite. High-field P-31-NMR spectroscopy showed that PPi was more abundant than ATP in tachyzoites. Bisphosphonates (PPi analogues), drugs that are used in the treatment of bone diseases, inhibited proton transport and PPi hydrolysis in tachyzoite homogenates, and also inhibited intracellular proliferation of tachyzoites in tissue culture cells.

L16 ANSWER 10 OF 17 SCISEARCH

ACCESSION NUMBER: 2000:385352 SCISEARCH
THE GENUINE ARTICLE: 314FV
TITLE: AVP2, a sequence-divergent, K+-insensitive H+-translocating inorganic pyrophosphatase from arabidopsis
AUTHOR: Drozdowicz Y M; Kissinger J C; Rea P A (Reprint)
CORPORATE SOURCE: UNIV PENN, DEPT BIOL, INST PLANT SCI, PHILADELPHIA, PA 19104 (Reprint); UNIV PENN, DEPT BIOL, INST PLANT SCI, PHILADELPHIA, PA 19104
COUNTRY OF AUTHOR: USA
SOURCE: PLANT PHYSIOLOGY, (MAY 2000) Vol. 123, No. 1, pp. 353-362.
Publisher: AMER SOC PLANT PHYSIOLOGISTS, 15501 MONONA DRIVE, ROCKVILLE, MD 20855.
ISSN: 0032-0889.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE; AGRI
LANGUAGE: English
REFERENCE COUNT: 42
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Plant vacuolar H+-translocating inorganic pyrophosphatases (V-PPases; EC 3.6.1.1) have been considered to constitute a family of functionally and structurally monotonous intrinsic membrane **proteins**. Typified by AVP1 (V. Sarafian, Y. Kim, R.J. Poole, P.A. Rea [1992] Proc Natl Acad Sci USA 89: 1775-1779) from Arabidopsis, all characterized plant V-PPases share greater than 84% sequence identity and catalyze K+-**stimulated** H+ translocation. Here we describe the molecular and biochemical characterization of AVP2 (accession number AF182813), a sequence-divergent (36% identical) K+-insensitive, Ca2+-hypersensitive V-PPase active in both inorganic pyrophosphate hydrolysis and H+ translocation. The differences between AVP2 and AVP1 provide the first indication that plant V-PPases from the same organism fall into two distinct categories. Phylogenetic analyses of these and other V-PPase sequences extend this principle by showing that AVP2, rather than being an isoform of AVP1, is but one representative of a novel category of AVP2-like (type II) V-PPases that coexist with AVP1-like (type I) V-PPases not only in plants, but also in apicomplexan protists such as the malarial parasite **Plasmodium falciparum**.

L16 ANSWER 11 OF 17 SCISEARCH
DUPLICATE 3

ACCESSION NUMBER: 96:116360 SCISEARCH
THE GENUINE ARTICLE: TT475
TITLE: ISOLATION AND CHARACTERIZATION OF SEA SPONGE MYOSIN
AUTHOR: KANZAWA N (Reprint); TAKANOOHMURO H; MARUYAMA K
CORPORATE SOURCE: CHIBA UNIV, FAC SCI, DEPT BIOL, CHIBA 263, JAPAN (Reprint); UNIV TOKYO, FAC MED, DEPT PHARMACOL, TOKYO 113, JAPAN

THIS PAGE BLANK (USPTO)

COUNTRY OF AUTHOR: JAPAN
SOURCE: ZOOLOGICAL SCIENCE, (DEC 1995) Vol. 12, No. 6, pp. 765-769.
ISSN: 0289-0003.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: LIFE; AGRI
LANGUAGE: ENGLISH
REFERENCE COUNT: 17
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB Myosin was purified to a homogeneity from sea sponge, Halichondria okadai. The myosin consisted of 220 kDa heavy chain, 18 kDa calcium binding light chain and 21 kDa phosphorylatable light chain. Rotary shadowed images showed the two headed myosin (myosin II) with a 160 nm tail. The myosin was less soluble in a KCl solution as compared to rabbit skeletal myosin.

The K+-**stimulated** and Ca2+-**stimulated** **ATPase** activities of sea sponge myosin were 0.46 and 0.07 mu mol Pi min(-1) mg(-1), respectively. The Mg2+ activated myosin **ATPase** activity showed no significant enhancement by the addition of rabbit skeletal muscle actin despite that the light chain was phosphorylated by myosin light chain kinase from chicken gizzard. Sea sponge myosin 18 kDa light chain bound to Ca2+ ion but was not phosphorylated like Physarum **plasmodia** myosin light chains.

L16 ANSWER 12 OF 17 SCISEARCH
ACCESSION NUMBER: 95:282492 SCISEARCH
THE GENUINE ARTICLE: QT916
TITLE: DRUG-**STIMULATED ATPASE** ACTIVITY OF THE HUMAN P-GLYCOPROTEIN

AUTHOR: SCARBOROUGH G A (Reprint)
CORPORATE SOURCE: UNIV N CAROLINA, DEPT PHARMACOL, CHAPEL HILL, NC, 27599 (Reprint)
COUNTRY OF AUTHOR: USA

SOURCE: JOURNAL OF BIOENERGETICS AND BIOMEMBRANES, (FEB 1995) Vol. 27, No. 1, pp. 37-41.
ISSN: 0145-479X.
DOCUMENT TYPE: General Review; Journal
FILE SEGMENT: LIFE
LANGUAGE: ENGLISH
REFERENCE COUNT: 41
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB The human multidrug resistance **protein**, or P-glycoprotein (Pgp), exhibits a high-capacity drug-dependent ATP hydrolytic activity that is a direct reflection of its drug transport capability. This activity is readily measured in membranes isolated from cultured insect cells infected with a baculovirus carrying the human mdrl cDNA. The drug-**stimulated** **ATPase** activity is a useful alternative to conventional screening systems for identifying high-affinity drug substrates of the Pgp with potential clinical value as chemosensitizers for tumor cells that have become drug resistant. Using this assay system, a variety of drugs have been directly shown to interact with the Pgp. Many of the drugs **stimulate** the Pgp **ATPase** activity, but certain drugs bind tightly to the drug-binding site of the Pgp without eliciting ATP hydrolysis. Either class of drugs may be useful as chemosensitizing agents. The baculovirus/insect cell Pgp **ATPase** assay system may also facilitate future studies

of the molecular structure and mechanism of the Pgp.

L16 ANSWER 13 OF 17 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
ACCESSION NUMBER: 1993:346268 BIOSIS
DOCUMENT NUMBER: PREV199396043268
TITLE: Absence of transitory (Ca-2+)-i flux during early in vitro metacyclogenesis of Trypanosoma cruzi.
AUTHOR(S): Krassner, Stuart M. (1); Chang, Johnny; Pak, Sung; Luc, Kim-Oanh; Granger, Barbara
CORPORATE SOURCE: (1) Dep. Dev. Cell Biol., Univ. Calif., Irvine, CA 92717 USA
SOURCE: Journal of Eukaryotic Microbiology, (1993) Vol. 40, No. 2, pp. 224-230.
ISSN: 1066-5234.
DOCUMENT TYPE: Article
LANGUAGE: English
AB The phorbol ester TPA (phorbol 12-myristate 13-acetate) substitutes for CO-2 as an agonist for transforming Trypanosoma cruzi epimastigotes to the metacyclic trypomastigote stage in a starvation medium consisting of phosphate buffered saline + 10 mM proline, 10 mM sodium acetate and 0.035% NaHCO-3. Since TPA is thought to **stimulate protein** kinase C by mimicking the activity of the secondary messenger diacylglycerol, the above result suggested that T. cruzi metacyclogenesis could be activated by a Ca-2+-dependent **protein** kinase C signal induction pathway. Accordingly, cytosolic calcium Bux ((Ca-2+)-i) in epimastigotes, activated with 5% CO-2 or TPA (10-7 M), was measured with the Ca-2+ molecular probe, fluo-3AM. In addition, (Ca-2+)-i was measured in cells incubated with putative metacyclogenic agonists (e.g. proline, glutamate, bioamines, ionophores and catecholamines). None of the compounds studied, except for EGTA, affected cytosolic Ca-2+ levels. Control assays with 11 mu-M thapsigargin, which mobilizes noncytoplasmic Ca-2+ stores by inhibiting endoplasmic reticulum Ca-2+-**ATPase,** validated our fluorometric assay procedure. Although thapsigargin significantly increases cytoplasmic Ca-2+ fluorescence, it has no effect on transformation. The **protein** kinase C inhibitors staurosporine, H-7 and HA 1004 were tested for their effect on T. cruzi metacyclogenesis. Low concentrations of staurosporine and HA 1004 significantly elevated Peru strain transformation while H-7 had no effect on Peru strain metacyclogenesis. Inhibitor H-7 did significantly depress CL transformation. The results indicate that induction of T. cruzi metacyclic trypomastigote formation by CO-2 and TPA is not accompanied by changes in cytosolic Ca-2+ and do not provide supporting evidence for participation of a **protein** kinase C-mediated phosphoinositide cascade in metacyclogenesis.

L16 ANSWER 14 OF 17 MEDLINE on STN DUPLICATE 4
ACCESSION NUMBER: 92042085 MEDLINE
DOCUMENT NUMBER: 92042085 PubMed ID: 1834662
TITLE: **Stimulation** of the interaction between actin and myosin by Physarum caldesmon-like **protein** and smooth muscle caldesmon.
AUTHOR: Ishikawa R; Okagaki T; Higashi-Fujime S; Kohama K
CORPORATE SOURCE: Department of Pharmacology, Gunma University School of Medicine, Japan.
SOURCE: JOURNAL OF BIOLOGICAL CHEMISTRY, (1991 Nov 15) 266

```
                    (32) 21784-90.
                    Journal code: 2985121R. ISSN: 0021-9258.
PUB. COUNTRY:       United States
DOCUMENT TYPE:      Journal; Article; (JOURNAL ARTICLE)
LANGUAGE:           English
FILE SEGMENT:       Priority Journals
ENTRY MONTH:        199112
ENTRY DATE:         Entered STN: 19920124
                    Last Updated on STN: 19970203
                    Entered Medline: 19911223
```

AB We have purified an actin-binding **protein** from the **plasmodia** of a lower eukaryote, Physarum polycephalum, with an apparent molecular mass of 210,000 daltons on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. This **protein** bound to actin filaments with a stoichiometry of 1:7-8 in a Ca(2+)-calmodulin-dependent manner. Antibody raised against caldesmon from smooth muscle cross-reacted with the 210-kDa **protein**. In vitro motility assay revealed that the 210-kDa **protein** increased the sliding velocity of actin filaments on Physarum myosin. The 210-kDa **protein** more than doubled the actin-activated **ATPase** activity of Physarum myosin under comparative conditions of in vitro motility assay. Further increases in the concentration of the 210-kDa **protein** decreased its **stimulatory** effects. Ca(2+)-calmodulin prevented the **stimulatory** effects of the 210-kDa **protein**. Unexpectedly, smooth muscle caldesmon also increased the sliding velocity of actin filaments on smooth muscle myosin at lower concentrations. The well-known inhibitory effect of smooth muscle caldesmon on the actin-myosin interaction was observed with this motility assay when the concentration of the caldesmon was increased further. The **stimulatory** and inhibitory effects were confirmed by measurements of actin-activated **ATPase** activity of smooth muscle myosin. From estimations of the intracellular concentrations of the 210-kDa **protein** and smooth muscle caldesmon in vivo, it appears that effects of the former and the latter on actin-myosin interactions in vivo are **stimulatory** and inhibitory, respectively.

```
L16  ANSWER 15 OF 17  SCISEARCH  COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER:      91:631866  SCISEARCH
THE GENUINE ARTICLE: GP804
TITLE:                 STIMULATION OF THE INTERACTION BETWEEN
                       ACTIN AND MYOSIN BY PHYSARUM CALDESMON-LIKE
                       PROTEIN AND SMOOTH-MUSCLE CALDESMON
AUTHOR:                ISHIKAWA R (Reprint); OKAGAKI T; HIGASHIFUJIME S;
                       KOHAMA K
CORPORATE SOURCE:      GUNMA UNIV, SCH MED, DEPT PHARMACOL, MAEBASHI, GUNMA
                       371, JAPAN (Reprint); NIHON UNIV, PHYS SCI LABS,
                       FUNABASHI, CHIBA 274, JAPAN; NAGOYA UNIV, FAC SCI,
                       DEPT MOLEC BIOL, NAGOYA, AICHI 464, JAPAN
COUNTRY OF AUTHOR:     JAPAN
SOURCE:                JOURNAL OF BIOLOGICAL CHEMISTRY, (1991) Vol. 266,
                       No. 32, pp. 21784-21790.
DOCUMENT TYPE:         Article; Journal
FILE SEGMENT:          LIFE
LANGUAGE:              ENGLISH
REFERENCE COUNT:       30
                      *ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS*
```

AB We have purified an actin-binding **protein** from the **plasmodia** of a lower eukaryote, Physarum polycephalum, with an apparent molecular mass of 210,000 daltons on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. This **protein** bound to actin filaments with a stoichiometry of 1:7-8 in a Ca2+-calmodulin-dependent manner. Antibody raised against caldesmon from smooth muscle cross-reacted with the 210-kDa **protein**. In vitro motility assay revealed that the 210-kDa **protein** increased the sliding velocity of actin filaments on Physarum myosin. The 210-kDa **protein** more than doubled the actin-activated **ATPase** activity of Physarum myosin under comparative conditions of in vitro motility assay. Further increases in the concentration of the 210-kDa **protein** decreased its **stimulatory** effects. Ca2+-calmodulin prevented the **stimulatory** effects of the 210-kDa **protein**. Unexpectedly, smooth muscle caldesmon also increased the sliding velocity of actin filaments on smooth muscle myosin at lower concentrations. The well-known inhibitory effect of smooth muscle caldesmon on the actin-myosin interaction was observed with this motility assay when the concentration of the caldesmon was increased further. The **stimulatory** and inhibitory effects were confirmed by measurements of actin-activated **ATPase** activity of smooth muscle myosin. From estimations of the intracellular concentrations of the 210-kDa **protein** and smooth muscle caldesmon in vivo, it appears that effects of the former and the latter on actin-myosin interactions in vivo are **stimulatory** and inhibitory, respectively.

L16 ANSWER 16 OF 17 MEDLINE on STN DUPLICATE 5

ACCESSION NUMBER: 89040035 MEDLINE
DOCUMENT NUMBER: 89040035 PubMed ID: 2972931
TITLE: Purification of **Plasmodium falciparum** digestive vacuoles and partial characterization of the vacuolar membrane **ATPase**.
AUTHOR: Choi I; Mego J L
CORPORATE SOURCE: Department of Biology, University of Alabama, Tuscaloosa 35487.
SOURCE: MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (1988 Oct) 31 (1) 71-8.
Journal code: 8006324. ISSN: 0166-6851.
PUB. COUNTRY: Netherlands
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 198812
ENTRY DATE: Entered STN: 19900308
Last Updated on STN: 19900308
Entered Medline: 19881220

AB **Plasmodium falciparum** digestive vacuoles containing ferric oxide granules were purified from parasite homogenates by centrifugation on discontinuous sucrose gradients. Digestive vacuole membranes prepared by osmotic lysis and washed with KCl showed no detectable contamination by erythrocyte membrane **proteins** and only minimal contamination by non-vacuolar parasite **proteins**. Purified vacuolar membranes were 2.6-fold enriched in total parasite membrane **ATPase** activity. This **ATPase** was optimally active at pH 7 in the

presence of at least 2 mM Mg2+. Ca2+ and Mn2+ were approximately 80-90% as effective as Mg2+, and Zn2+, Co2+ and Fe2+ also exerted some **stimulatory** effect. The vacuolar membrane also hydrolyzed GTP, UTP, CTP and ADP, but AMP and 3',5'-cyclic AMP were hydrolyzed only one-tenth as effectively as ATP. The **ATPase** was unaffected by vanadate, ouabain or oligomycin but was significantly inhibited by the proton pump inhibitors NEM and NBD-Cl. Of 6 antimalarial drugs tested, quinine and quinacrine were the most effective inhibitors and mefloquine was the least effective.

L16 ANSWER 17 OF 17 EMBASE on STN
ACCESSION NUMBER: 74138989 EMBASE
DOCUMENT NUMBER: 1974138989
TITLE: The biosynthesis of plasmodial myosin during starvation of Physarum polycephalum.
AUTHOR: White F.H.; Lascelles J.
CORPORATE SOURCE: Dept. Bacteriol., Univ. California, Los Angeles, Calif. 90025, United States
SOURCE: Biochemical Journal, (1973) 135/4 (639-647). CODEN: BIJOAK
DOCUMENT TYPE: Journal
FILE SEGMENT: 029 Clinical Biochemistry
LANGUAGE: English
AB The actomyosin **protein** complex of Physarum polycephalum was prepared from vegetative and starved **plasmodia.** The yield of actomyosin per unit wet wt was the same from both types of **plasmodia.** Myosin was resolved from the complex by gel filtration and purified by ion exchange chromatography. The Ca2+ **stimulated adenosine triphosphatase** activities of myosin preparations from vegetative and starved **plasmodia** were not appreciably different. Synthesis of myosin de novo was shown to occur during the starvation phase of the life cycle by the isolation of labelled myosin preparations from **plasmodia** starved in the presence of [2 14C]glycine. Fractionation of polyacrylamide gels after gel filtration of labelled myosin confirmed the presence of label in the **adenosine triphosphatase** active myosin band. It is concluded that during starvation myosin synthesis continues although there is a net loss of approx. 50% of the total **protein.** Sodium dodecyl sulphate polyacrylamide gel electrophoresis of Physarum myosin showed the presence of low molecular weight components of the molecule, similar to those of muscle myosins. The content and composition of the free amino acid pool to Physarum was measured at various time intervals during the vegetative and starvation phases of the life cycle.

(FILE 'MEDLINE' ENTERED AT 14:43:41 ON 07 OCT 2003)
L29 12195 SEA FILE=MEDLINE ABB=ON PLU=ON "PLASMODIUM FALCIPARUM"/CT
L30 28410 SEA FILE=MEDLINE ABB=ON PLU=ON ADENOSINETRIPHOSPHATASE/CT
L31 17 SEA FILE=MEDLINE ABB=ON PLU=ON L29 AND L30

L31 ANSWER 1 OF 17 MEDLINE on STN
AN 2002675294 MEDLINE
TI Transport processes in Plasmodium falciparum-infected erythrocytes:

potential as new drug targets.
AU Krishna Sanjeev; Eckstein-Ludwig Ursula; Joet Thierry; Uhlemann Anne-Catrin; Morin Christophe; Webb Richard; Woodrow Charles; Kun Jurgen F J; Kremsner Peter G
SO INTERNATIONAL JOURNAL FOR PARASITOLOGY, (2002 Dec 4) 32 (13) 1567-73. Ref: 39
Journal code: 0314024. ISSN: 0020-7519.
AB Plasmodium falciparum infection induces alterations in the transport properties of infected erythrocytes that have recently been defined using electrophysiological techniques. Mechanisms responsible for transport of substrates into intraerythrocytic parasites have also been clarified by studies of three substrate-specific (hexose, nucleoside and aquaglyceroporin) parasite plasma membrane transporters. These have been characterised functionally using the Xenopus laevis oocyte heterologous expression system. The same expression system is currently being used to define the function of parasite 'P' type ATPases responsible for intraparasitic [Ca(2+)] homeostasis. We review studies on these transport processes and examine their potential as novel drug targets.

L31 ANSWER 2 OF 17 MEDLINE on STN
AN 2001479314 MEDLINE
TI Characterization of P-type ATPase 3 in Plasmodium falciparum.
AU Rozmajzl P J; Kimura M; Woodrow C J; Krishna S; Meade J C
SO MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (2001 Sep 3) 116 (2) 117-26.
Journal code: 8006324. ISSN: 0166-6851.
AB We report the nucleotide sequence, derived amino acid sequence and expression profile of P-type ATPase 3 (PfATPase3) from Plasmodium falciparum. An open reading frame of 7362 nucleotides, interrupted by a single intron of 168 nt, encoded a protein product of 2394 amino acids with a predicted MW of 282791 Da. Hydropathy analysis of PfATPase3 revealed six amino-terminal and six carboxyl-terminal membrane spanning regions (M1-12) flanking a large hydrophilic domain with a smaller hydrophilic loop between M4 and M5. Based on a phylogenetic comparison of conserved domains present in P-type ATPases from other organisms, PfATPase3 resembled a Type-V ATPase for which the transport affinity is unknown. The PfATPase3 topology was interrupted by four regions, termed 'inserts', unique to malarial P-type ATPases, which were high in asparagine residues and charged amino acids (inserts I1-I4). Inserts I1 and I3 also contained repeated amino acid motifs. The number and composition of repeated amino acid motifs in insert I3 were variable in seven P. falciparum strains tested. PfATPase3 was 80.2% similar to the non-insert portions of P. yoelii ATPase3, although their inserts differed in length and composition. PfATPase3 mRNA was most abundant relative to beta-tubulin during the latter half of the erythrocytic cycle and was also present in gametocytes. Using affinity-purified antibody to a 14 amino acid PfATPase3 epitope, a 260 kDa protein was detected by Western analysis. Based on immunofluorescence, the PfATPase3 protein was located intracellularly in gametocytes and, to a lesser extent, in late erythrocytic stages.

L31 ANSWER 3 OF 17 MEDLINE on STN
AN 2000386953 MEDLINE
TI Molecular cloning of a gene encoding a 20S proteasome beta subunit from Plasmodium falciparum.
AU Li G D; Li J L; Mugthin M; Ward S A

THIS PAGE BLANK (USPTO)

SO INTERNATIONAL JOURNAL FOR PARASITOLOGY, (2000 May) 30 (6) 729-33.
Journal code: 0314024. ISSN: 0020-7519.
AB A novel gene was cloned from Plasmodium falciparum. Database searches indicated this gene to be a member of the 20S proteasome beta-subunit family. Comparison of the gene's genomic DNA sequence with cDNA sequence revealed a 156-bp intron 85 bp downstream from the start codon. The nucleotide sequence of the gene contains one open reading frame encoding 265 amino acids with a predicted molecular mass of 30.9 kDa and a pI of 6.2. Northern blot analysis showed the transcript size to be approximately 1.6 kb indicating that some 800 bp of the transcript is non-coding.

L31 ANSWER 4 OF 17 MEDLINE on STN
AN 2000002518 MEDLINE
TI Cloning and partial characterization of the proteasome S4 ATPase from Plasmodium falciparum.
AU Certad G; Abrahem A; Georges E
SO EXPERIMENTAL PARASITOLOGY, (1999 Nov) 93 (3) 123-31.
Journal code: 0370713. ISSN: 0014-4894.
AB Certad, G., Abrahem, A., and Georges, E. 1999. Cloning and Partial characterization of the proteasome S4 ATPase from Plasmodium falciparum. Experimental Parasitology 93, 123-131. The ATP-ubiquitin-proteasome pathway mediates the nonlysosomal degradation of cytosolic proteins in eukaryotic cells. The activities of this pathway have been shown to regulate cell growth and differentiation through modulation of regulatory proteins. The proteasome is a large complex consisting of two multisubunit structures, the 20S and 19S(PA700) or P28 complexes, that combine to form the 26S particles. In this study, we describe the cloning of a cDNA encoding the proteasome subunit 4 ATPase homologue from Plasmodium falciparum (PFS4). Analysis of the PFS4 cDNA sequence shows an open reading frame encoding a deduced protein of 455 amino acids. Moreover, comparison of PFS4 cDNA sequence to that of genomic fragments encoding PFS4 showed identical sequences with no detectable introns. Database searches revealed a high sequence identity to those of rice, yeast, mouse, Drosophila, and human S4 ATPases. However, PFS4 contains two unique inserts of nine and seven amino acid residues in the N-terminal domain. Interestingly, only the rice S4 contains the latter (seven amino acids) insert with four identical amino acids. In vitro expression of the full-length cDNA encoding the PFS4, using a transcription-translation-coupled reticulocyte lysate, shows a 50-kDa [(35)S]methionine-labeled protein which was immunoprecipitated with PFS4 anti-peptide antiserum. Southern blot analysis of genomic DNA digests shows a single gene copy of PFS4 in P. falciparum. Of interest was the effect of the proteasome-specific natural product, lactacystin, on the growth of the parasite, with IC(50) values of 0.6-0.92 microM. The latter IC(50) values of lactacystin for different clones of P. falciparum are comparable to those obtained for mammalian cell lines (0.65 microM), suggesting the presence of a conserved proteasome complex. Moreover, lactacystin was equally toxic to drug-sensitive and resistant parasites.
Copyright 1999 Academic Press.

L31 ANSWER 5 OF 17 MEDLINE on STN
AN 97418800 MEDLINE
TI A Plasmodium falciparum homologue of the ATPase subunit of a multi-protein complex involved in chromatin remodelling for

transcription.
AU Ji D D; Arnot D E
SO MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (1997 Sep) 88 (1-2) 151-62.
Journal code: 8006324. ISSN: 0166-6851.
AB A Plasmodium falciparum homologue of one of the components of a chromatin-remodelling complex which controls binding of transcription factors to nucleosome core particles has been cloned and characterised. The gene encodes 1422 amino acids with an estimated molecular mass of 167 kDa. The protein, SNF2L, shares 60% amino acid identity in its conserved DNA-dependent ATPase domain with yeast transcription factors originally identified by characterising mating type switch mutants. It also contains sequences related to the so-called SWI3, ADA2, N-CoR and TFIIIB B" or SANT DNA binding domains which are characteristic of these transcriptional activation factors. The SNF2L gene has two short introns in the 3' region of the coding sequence of the gene and is transcribed into a single approximately 6.5 kb messenger RNA species which is present throughout the asexual stages of the cell cycle. Southern blotting and pulsed field gel electrophoresis experiments show that SNF2L is a single copy gene. located on P. falciparum chromosome 11.

L31 ANSWER 6 OF 17 MEDLINE on STN
AN 96408665 MEDLINE
TI Analysis of a cation-transporting ATPase of Plasmodium falciparum.
AU Dyer M; Jackson M; McWhinney C; Zhao G; Mikkelsen R
SO MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (1996 Jun) 78 (1-2) 1-12.
Journal code: 8006324. ISSN: 0166-6851.
AB We have cloned and characterised one gene, PfATPase4 which encodes a P-type ATPase containing all the primary sequence motifs characteristic of this class of transmembrane ion transporters, and also a fragment of a second P. falciparum P-type ATPase pseudogene (PfATPase5). Analysis of conserved domains and motifs of specific ATPases reveals that PfATPase4 is most analogous to Ca2+ ATPases of the endoplasmic reticulum. The PfATPase4 gene gives rise to a transcript of 8 kb shortly after erythrocyte invasion. Although this mRNA is not detected in later stages, the protein detected immunologically at 190 kDa persists throughout and is detected in free merozoites. Immunofluorescence microscopy reveals that the PfATPase4 protein is concentrated in discrete compartments at the periphery of the parasite. Detailed sequence and structural analyses of these and the other P-type ATPases of P. falciparum described previously, reveals that they comprise an unusual family in several respects. Firstly, the large number of non-homologous genes so far characterised reflects the complexities of ionic regulation in the diverse environments encountered by the parasite. Secondly, the plasmodial P-type ATPase family may be classified both at primary sequence and structural levels into two distinct groups-those typical of P-type ATPases (including PfATPase4) and those which are much more divergent. A third complexity is illustrated by the fact that one of the other members [1] here termed PfATPase6, has an even greater similarity to the sarcoplasmic reticulum Ca2+ ATPases than does PfATPase4, which raises questions about the possible functional relationship between these two members.

L31 ANSWER 7 OF 17 MEDLINE on STN
AN 95309711 MEDLINE

TI Cloning of a new cation ATPase from Plasmodium falciparum: conservation of critical amino acids involved in calcium binding in mammalian organellar Ca(2+)-ATPases.
AU Trottein F; Thompson J; Cowman A F
SO GENE, (1995 May 26) 158 (1) 133-7.
Journal code: 7706761. ISSN: 0378-1119.
AB In order to study molecules that may be involved in pH gradient formation in Plasmodium, we have identified a novel cation-translocating ATPase (P-type ATPase) gene from P. falciparum (Pf). We report the full-length nucleotide and deduced amino acid (aa) sequences of this gene that we called PfATPase4. The PfATPase4 protein shares features with the different members of eukaryotic P-type ATPases, such as a similar transmembrane (TM) organization and aa identity in functionally important regions. Interestingly, the PfATPase4 protein possesses conserved aa involved in calcium binding in mammalian organellar Ca(2+)-ATPases.

L31 ANSWER 8 OF 17 MEDLINE on STN
AN 95154293 MEDLINE
TI Molecular cloning and sequence of two novel P-type adenosinetriphosphatases from Plasmodium falciparum.
AU Trottein F; Cowman A F
SO EUROPEAN JOURNAL OF BIOCHEMISTRY, (1995 Jan 15) 227 (1-2) 214-25.
Journal code: 0107600. ISSN: 0014-2956.
AB We have identified two novel P-type ATPase genes from Plasmodium falciparum and report the full-length nucleotide and derived amino acid sequence of the ATPase2 gene from P. falciparum (PfATPase2). PfATPase2 is phylogenetically remote from the different members of prokaryotic and mammalian P-type ATPases but shares features with a putative membrane-spanning Ca2+ ATPase involved in ribosome function in yeast. PfATPase2 is expressed during the intraerythrocytic life cycle of the parasite and appears to be required in the late stages of its asexual development. We also present the partial sequence of another malarial gene displaying sequence similarity with the family of P-type transporting ATPases (PfATPase4). We have analysed the organisation of the genes encoding the P-type ATPases of P. falciparum and show that they are a highly dispersed gene family.

L31 ANSWER 9 OF 17 MEDLINE on STN
AN 95021518 MEDLINE
TI Cloning and characterization of the vacuolar ATPase B subunit from Plasmodium falciparum.
AU Karcz S R; Herrmann V R; Trottein F; Cowman A F
SO MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (1994 May) 65 (1) 123-33.
Journal code: 8006324. ISSN: 0166-6851.
AB The transvacuolar pH gradient determines, to a significant extent, the distribution of the antimalarial drug chloroquine in Plasmodium falciparum. A proton pump, similar to the vacuolar ATPase found in many cell types, appears to regulate a pH gradient across the membranes of acidic compartments of the parasite. In order to understand and define the components involved in the maintenance of the vacuolar pH gradient, we have cloned and characterized a gene, designated VAP B, encoding a P. falciparum homologue of the B subunit of the vacuolar ATPase. The VAP B gene encodes a protein of 494 amino acids which has between 69% and 74% amino acid identity with the sequences of vacuolar ATPase B subunits of other organisms. The VAP B gene exists as a single copy gene on chromosome 4 that gives rise to a RNA transcript of 2.4 kb. Antibodies raised to the

VAP B protein react specifically with a protein of 56-kDa, consistent with the size predicted from the gene sequence and with the homologous protein from other organisms. The 56-kDa protein is expressed throughout the asexual life cycle and subcellular localization by indirect immunofluorescence shows that the protein has a heterogeneous distribution over most of the parasite. This suggests that the function of the vacuolar proton ATPase is not confined to the regulation of the pH of the digestive vacuole.

L31 ANSWER 10 OF 17 MEDLINE on STN
AN 94131050 MEDLINE
TI Plasmodium falciparum: further characterization of putative cation ATPases.
AU Krishna S; Cowan G M; Robson K J; Meade J C
SO EXPERIMENTAL PARASITOLOGY, (1994 Feb) 78 (1) 113-7.
Journal code: 0370713. ISSN: 0014-4894.

L31 ANSWER 11 OF 17 MEDLINE on STN
AN 93241225 MEDLINE
TI Cloning and characterization of a vacuolar ATPase A subunit homologue from Plasmodium falciparum.
AU Karcz S R; Herrmann V R; Cowman A F
SO MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (1993 Apr) 58 (2) 333-44.
Journal code: 8006324. ISSN: 0166-6851.
AB The distribution of the antimalarial drug chloroquine is determined to a significant extent by a transvacuolar pH gradient in Plasmodium falciparum. A proton pump similar to the vacuolar ATPase found in many cell types has been suggested to maintain a pH gradient across the membranes of acidic compartments in the parasite. In order to understand and define the components involved in the mechanism of acidification of parasite vesicles, we have cloned and characterized a gene, designated VAP-A, encoding a P. falciparum homologue of the catalytic A subunit of the vacuolar ATPase. The VAP-A gene encodes a polypeptide of 611 amino acids which shows between 56 to 61% amino acid identity over its entire length with the sequences of vacuolar ATPase A subunits from several species. The VAP-A gene exists as a single copy gene on P. falciparum chromosome 13 and gives rise to a transcript of 3.7 kb. Antibodies raised against a VAP-A gene segment expressed in Escherichia coli react specifically with a 67-kDa polypeptide, consistent with the size predicted from the sequence and with the size of the corresponding polypeptide in other organisms. The 67-kDa protein is present throughout the asexual erythrocytic cycle and is expressed at similar levels in 5 P. falciparum isolates of differing chloroquine sensitivity. Sequence analysis of the coding region of the VAP-A gene from 2 chloroquine-sensitive and 3 chloroquine-resistant isolates has shown no changes that are linked to chloroquine resistance. Therefore, a proposed chloroquine resistance-linked vacuolar acidification defect does not involve mutations in the VAP-A gene in the isolates we have studied.

L31 ANSWER 12 OF 17 MEDLINE on STN
AN 93132070 MEDLINE
TI A family of cation ATPase-like molecules from Plasmodium falciparum.
AU Krishna S; Cowan G; Meade J C; Wells R A; Stringer J R; Robson K J
SO JOURNAL OF CELL BIOLOGY, (1993 Jan) 120 (2) 385-98.
Journal code: 0375356. ISSN: 0021-9525.
AB We report the nucleotide and derived amino acid sequence of the

ATPase 1 gene from Plasmodium falciparum. The amino acid sequence shares homology with the family of "P"-type cation translocating ATPases in conserved regions important for nucleotide binding, conformational change, or phosphorylation. The gene, which is present on chromosome 5, has a product longer than any other reported for a P-type ATPase. Interstrain analysis from 12 parasite isolates by the polymerase chain reaction reveals that a 330-bp nucleotide sequence encoding three cytoplasmic regions conserved in cation ATPases (regions a-c) is of constant length. By contrast, another 360-bp sequence which is one of four regions we refer to as "inserts" contains arrays of tandem repeats which show length variation between different parasite isolates. Polymorphism results from differences in the number and types of repeat motif contained in this insert. Inserts are divergent in sequence from other P-type ATPases and share features in common with many malarial antigens. Studies using RNA from the erythrocytic stages of the malarial life cycle suggest that ATPase 1 (including the sequence which encodes tandem repeats) is expressed at the large ring stage of development. Immunolocalization has identified ATPase 1 to be in the region of the parasite plasma membrane and pigment body. These findings suggest a possible model for the genesis of malarial antigens.

L31 ANSWER 13 OF 17 MEDLINE on STN
AN 92134363 MEDLINE
TI Energy dependence of chloroquine accumulation and chloroquine efflux in Plasmodium falciparum.
AU Krogstad D J; Gluzman I Y; Herwaldt B L; Schlesinger P H; Wellems T E
SO BIOCHEMICAL PHARMACOLOGY, (1992 Jan 9) 43 (1) 57-62.
Journal code: 0101032. ISSN: 0006-2952.
AB Chloroquine inhibits the growth of susceptible malaria parasites at low (nanomolar) concentrations because of an energy-requiring drug-concentrating mechanism in the parasite secondary lysosome (food vacuole) which is dependent on the acidification of that vesicle. Chloroquine resistance results from another energy-requiring process: efflux of chloroquine from the resistant parasite with a half-time of 2 min. Chloroquine efflux is inhibited reversibly by the removal of metabolizable substrate (glucose); it is also reduced by the ATPase inhibitor vanadate. These results suggest that chloroquine efflux is an energy-requiring process dependent on the generation and hydrolysis of ATP. Chloroquine efflux cannot be explained by differences in drug accumulation between chloroquine-susceptible and -resistant parasites because the 40-50-fold difference in initial efflux rates between -susceptible and -resistant parasites is unchanged when both parasites contain the same amount of chloroquine. Although chloroquine efflux is phenotypically similar to the efflux of anticancer drugs from multidrug-resistant (mdr) mammalian cells, it is not linked to either of the mdr-like genes of the parasite.

L31 ANSWER 14 OF 17 MEDLINE on STN
AN 91101662 MEDLINE
TI Accumulation of chloroquine by membrane preparations from Plasmodium falciparum.
AU Herwaldt B L; Schlesinger P H; Krogstad D J
SO MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (1990 Sep-Oct) 42 (2) 257-67.
Journal code: 8006324. ISSN: 0166-6851.

AB Chloroquine susceptibility and resistance have been associated respectively with the uptake and efflux of chloroquine by Plasmodium falciparum. We made membrane preparations from parasitized and unparasitized red cells in order to study chloroquine accumulation in a cell-free system. The accumulation of [3H]chloroquine by these preparations is inhibited by unlabeled chloroquine and thus is specific. Only membranes from parasitized red cells demonstrate time-dependent chloroquine accumulation; membranes from unparasitized red cells do not. Chloroquine accumulation is eliminated by detergent (0.05% Triton X-100) and reduced by a hypertonic medium, consistent with accumulation inside membrane vesicles rather than binding to membranes. Accumulation is energy dependent; it has a specific requirement for ATP, which cannot be replaced with GTP, CTP, UTP, TTP or ADP, an apparent Km of 21 microM and an apparent Vmax of 4.6 pmol (mg protein)-1 h-1. Vesicle acidification is MgATP dependent, and is reversed by NH4Cl. Chloroquine accumulation is inhibited by reduced medium pH, N-ethylmaleimide or oligomycin, but not by vanadate or ouabain. These studies demonstrate that membrane vesicles prepared from parasitized red cells provide a model system for the study of chloroquine accumulation by P. falciparum.

L31 ANSWER 15 OF 17 MEDLINE on STN
AN 91077466 MEDLINE
TI Ion metabolism in malaria-infected erythrocytes.
AU Tanabe K
SO BLOOD CELLS, (1990) 16 (2-3) 437-49. Ref: 61
Journal code: 7513567. ISSN: 0340-4684.
AB Malaria parasites of the genus Plasmodium spend much of their asexual life cycle inside the erythrocytes of their vertebrate hosts. Parasites presumably have to exploit metabolic and transport mechanisms to adapt themselves to the host erythrocyte's physicochemical environment. This review surveys the metabolism and transport of Ca2+, alkali cations, and H+ in malaria-infected erythrocytes. The Ca2+ content of Plasmodium-infected erythrocytes increases as the parasite matures. An increase in the influx of extracellular Ca2+ into infected erythrocytes is evident at later stages of parasite development. In infected erythrocytes, Ca2+ is almost exclusively localized in the parasite compartment and changes but little in the cytosol of the host cell. The importance of Ca2+ in supporting the growth of intraerythrocytic parasites and the invasion of erythrocytes by the merozoite has been assessed by depletion of extracellular Ca2+ with chelators, or by disturbance of the metabolism and transport of Ca2+ with a variety of Ca2+ modulators. Membranes of malaria-infected erythrocytes change their permeability to alkali cations. Hence, levels of K+ decrease and levels of Na+ increase in the cytosol of infected erythrocytes. Intraerythrocytic parasites maintain a high K+, low Na+ state, suggesting a mechanism for transporting K+ inward and Na+ outward against concentration gradients of the alkali cations across the parasite plasma membrane and/or the parasitophorous vacuole membrane (PVM). Concomitantly, P. falciparum can grow in Na(+)-enriched human erythrocytes. Experimental evidence suggests that Plasmodium possesses in its plasma membrane a proton pump which is very sensitive to orthovanadate, carbonylcyanide m-chlorophenylhydrazone, a protonophore, and dicyclohexylcarbodiimide, an inhibitor of H(+)-ATPase, but is only slightly sensitive to inhibitors of bacterial and mitochondrial respiration, such as antimycin A, CN-,

or N3-, and ouabain, a Na+, K(+)-ATPase inhibitor. By operating this proton pump, parasites extrude H+ and thus generate an electrochemical gradient of protons (an internal negative membrane potential and a concentration gradient of protons) across the parasite plasma membrane. The electrochemical gradient apparently drives inward movement of Ca2+ and, possibly, glucose from the cytosol of infected erythrocytes. Little is known about the transport properties of the PVM. Recent sequence studies suggest that Plasmodium contains a cation-transporting ATPase which exhibits a high homology to the Ca2(+)-ATPase of rabbit skeletal muscle sarcoplasmic reticulum.(ABSTRACT TRUNCATED AT 250 WORDS)

L31 ANSWER 16 OF 17 MEDLINE on STN
AN 89040035 MEDLINE
TI Purification of Plasmodium falciparum digestive vacuoles and partial characterization of the vacuolar membrane ATPase.
AU Choi I; Mego J L
SO MOLECULAR AND BIOCHEMICAL PARASITOLOGY, (1988 Oct) 31 (1) 71-8. Journal code: 8006324. ISSN: 0166-6851.
AB Plasmodium falciparum digestive vacuoles containing ferric oxide granules were purified from parasite homogenates by centrifugation on discontinuous sucrose gradients. Digestive vacuole membranes prepared by osmotic lysis and washed with KCl showed no detectable contamination by erythrocyte membrane proteins and only minimal contamination by non-vacuolar parasite proteins. Purified vacuolar membranes were 2.6-fold enriched in total parasite membrane ATPase activity. This ATPase was optimally active at pH 7 in the presence of at least 2 mM Mg2+. Ca2+ and Mn2+ were approximately 80-90% as effective as Mg2+, and Zn2+, Co2+ and Fe2+ also exerted some stimulatory effect. The vacuolar membrane also hydrolyzed GTP, UTP, CTP and ADP, but AMP and 3',5'-cyclic AMP were hydrolyzed only one-tenth as effectively as ATP. The ATPase was unaffected by vanadate, ouabain or oligomycin but was significantly inhibited by the proton pump inhibitors NEM and NBD-Cl. Of 6 antimalarial drugs tested, quinine and quinacrine were the most effective inhibitors and mefloquine was the least effective.

L31 ANSWER 17 OF 17 MEDLINE on STN
AN 74085067 MEDLINE
TI Bongkrekic acid and the adenosinetriphosphate requirement of malaria parasites.
AU Trager W
SO EXPERIMENTAL PARASITOLOGY, (1973 Dec) 34 (3) 412-6. Journal code: 0370713. ISSN: 0014-4894.

FILE 'HCAPLUS' ENTERED AT 14:45:35 ON 07 OCT 2003
L32 8 SEA FILE=HCAPLUS ABB=ON PLU=ON (PLASMODIUM OR FALCIPARU M) AND ADENOSINETRIPHOSPHATASE
L33 0 SEA FILE=HCAPLUS ABB=ON PLU=ON L32 AND (MICROTUB? OR MICRO TUBUL? OR ((PROTEIN OR PEPTIDE OR POLYPROTEIN OR POLYPEPTIDE) AND STIMUL?))

(FILE 'HCAPLUS, MEDLINE, BIOSIS, EMBASE, WPIDS, CONFSCI, SCISEARCH, JICST-EPLUS, JAPIO' ENTERED AT 14:47:45 ON 07 OCT 2003)
L34 1 S L33
L35 0 S L34 NOT L15

FILE 'HOME' ENTERED AT 14:50:06 ON 07 OCT 2003

```
     (FILE 'HCAPLUS, MEDLINE, BIOSIS, EMBASE, WPIDS, CONFSCI, SCISEARCH,
     JICST-EPLUS, JAPIO' ENTERED AT 14:40:14 ON 07 OCT 2003)
L17          131 S "SAKOWICZ R"?/AU
L18          561 S "BERAUD C"?/AU                          - Author(s)
L19        10837 S "GUO J"?/AU
L20         4357 S "FREEDMAN R"?/AU
L21            3 S L17 AND L18 AND L19 AND L20
L22           42 S L17 AND (L18 OR L19 OR L20)
L23          108 S L18 AND (L19 OR L20)
L24            6 S L19 AND L20
L25        15730 S L17 OR L18 OR L19 OR L20
L26           14 S (PLASMODIUM OR FALCIPARUM) AND (L22 OR L23 OR L25)
L27           18 S L21 OR L24 OR L26
L28           10 DUP REM L27 (8 DUPLICATES REMOVED)
```

L28 ANSWER 1 OF 10 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 1

ACCESSION NUMBER: 2003:454460 HCAPLUS
DOCUMENT NUMBER: 139:31820
TITLE: **Plasmodium falciparum** kinesin motor protein KinI-1 with microtubule-stimulated ATPase activity and uses for diagnosis and treatment of malaria
INVENTOR(S): **Sakowicz, Roman; Beraud, Christophe; Guo, Jun; Freedman, Richard**
PATENT ASSIGNEE(S): Cytokinetics, Inc., USA
SOURCE: PCT Int. Appl., 68 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003048320	A2	20030612	WO 2002-US38360	20021127

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

US 2003104496	A1	20030605	US 2001-6780	20011130

PRIORITY APPLN. INFO.: US 2001-6780 A2 20011130
US 2002-86935 A2 20020228

AB The present invention relates to protein and cDNA sequences of a new kinesin motor protein, P. **falciparum** KinI-1 (PfKinI-1), and the use of these compns. for the diagnosis, treatment, or prevention of malaria. The invention provides isolated nucleic acid and amino acid sequences of kinesin superfamily motor protein KinI-1, which has microtubule stimulated ATPase activity and/or depolymerizes microtubules. The invention further relates to antibodies to PfKinI-1, methods of screening for PfKinI-1 modulators

using biol. active PfKinI-1, and kits for screening for PfKinI-1 modulators.

L28 ANSWER 2 OF 10 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2003:435206 HCAPLUS
DOCUMENT NUMBER: 139:909
TITLE: Sequences of **Plasmodium falciparum** kinesin KinI-1 and use for treating malaria
INVENTOR(S): **Sakowicz, Roman; Beraud, Christophe; Guo, Jun; Freedman, Richard**
PATENT ASSIGNEE(S): Cytokinetics, Inc., USA
SOURCE: U.S. Pat. Appl. Publ., 32 pp.
CODEN: USXXCO
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 2003104496	A1	20030605	US 2001-6780	20011130
WO 2003048320	A2	20030612	WO 2002-US38360	20021127

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: US 2001-6780 A2 20011130
US 2002-86935 A2 20020228

AB The invention provides sequences of **Plasmodium falciparum** kinesin KinI-1. The invention also relates to the use of KinI-1 for the diagnosis, treatment, or prevention of malaria.

L28 ANSWER 3 OF 10 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2003-492112 [46] WPIDS
CROSS REFERENCE: 2002-352124 [38]; 2003-491700 [46]
DOC. NO. CPI: C2003-131606
TITLE: New human microtubule motor protein (kinesin motor protein KinI-3) and nucleic acid, useful for diagnosing, preventing or treating e.g. cancer, restenosis, inflammation, neurological disorders or disorders of vesicular transport.
DERWENT CLASS: B04 D16
INVENTOR(S): BERAUD, C; DAVIES, K A; **FREEDMAN, R; GUO, J**; PATEL, U A
PATENT ASSIGNEE(S): (CYTO-N) CYTOKINETICS
COUNTRY COUNT: 1
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
US 2003036075	A1	20030220	(200346)*		31

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
US 2003036075 A1	CIP of	US 2000-675227	20000929
	Div ex	US 2001-967908	20010928
		US 2002-159151	20020531

FILING DETAILS:

PATENT NO	KIND	PATENT NO
US 2003036075 A1	CIP of	US 6461855

PRIORITY APPLN. INFO: US 2001-967908 20010928; US 2000-675227 20000929; US 2002-159151 20020531

AN 2003-492112 [46] WPIDS

CR 2002-352124 [38]; 2003-491700 [46]

AB US2003036075 A UPAB: 20030719

NOVELTY - An isolated microtubule motor protein, which has greater than 70 % sequence identity to a 1368 amino acid sequence (P1), given in the specification as measured using a sequence comparison algorithm, is new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) an isolated nucleic acid sequence:

(a) encoding the microtubule motor protein described above, where the motor the protein's activity includes microtubule depolymerization activity; or

(b) comprising a sequence which has greater than 60 % sequence identity with nucleotide comprising 6409 base pairs (N1), given in the specification;

(2) an expression vector comprising the nucleic acid encoding the microtubule motor protein;

(3) a host cell transfected with the vector of (2);

(4) screening for modulators of KinI-3 comprising:

(a) providing the biologically active KinI-3;

(b) contacting the biologically active KinI-3 with a candidate agent in a test and control concentration; and

(c) assaying for the level of KinI-3 activity, where the KinI-3 activity consists of binding activity or ATPase activity, and where a change in activity between the test and control concentration indicates a modulator; and

(5) a compound that modulates KinI-3, which is identified using the method of (4).

ACTIVITY - Cytostatic; Vasotropic; Immunomodulator; Antiinflammatory; Vulnerary; Antirheumatic; Antiarthritic; Antiarteriosclerotic; Antigout; Antipsoriatic; Antidiabetic; Ophthalmological; Immunosuppressive; Neuroprotective.

No biological data is given.

MECHANISM OF ACTION - ATPase Modulator; Microtubule Motor Protein Modulator.

USE - The KinI-3 protein, nucleic acid, or its modulator, is useful for diagnosing, preventing or treating cellular proliferation

(e.g. cancers (e.g. bronchogenic, carcinoma, Kaposi's sarcoma, lymphoma, leukemia, osteoid osteoma, glioblastoma, etc.), hyperplasia, restenosis, cardiac hypertrophy, immune disorders or inflammation), neurological disorders, or disorders of vesicular transport. These disorders also include atherosclerosis, hemangiomas, acoustic neuromas, vascular malfunctions, abnormal wound healing, rheumatoid arthritis, Bechet's disease, gout, psoriasis, diabetic retinopathy, corneal graft rejection, glaucoma, Osler Webber syndrome, etc. The protein, nucleic acid or the KinI-3 modulator regulates cell cycle, as well as cellular proliferation. The KinI-3 protein or nucleic acid is also useful for screening therapeutic agents or KinI-3 modulators, which may be used for treating the above-mentioned diseases or disorders.
Dwg.0/9

L28 ANSWER 4 OF 10 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 2
ACCESSION NUMBER: 2002:256423 HCAPLUS
DOCUMENT NUMBER: 136:274347
TITLE: Human kinesin motor protein KinI-3, protein and cDNA sequences and tissue expression
INVENTOR(S): Beraud, Christophe; **Guo, Jun**; **Freedman, Richard**; Patel, Umesh A.; Davies, Katherine A.
PATENT ASSIGNEE(S): Cytokinetics, Inc., USA
SOURCE: PCT Int. Appl., 68 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002026929	A2	20020404	WO 2001-US30750	20010928
WO 2002026929	A3	20020613		

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

US 6461855	B1	20021008	US 2000-675227	20000929
US 6432659	B1	20020813	US 2000-724510	20001127
US 6436686	B1	20020820	US 2000-723216	20001127
EP 1330535	A2	20030730	EP 2001-977340	20010928

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR

PRIORITY APPLN. INFO.: US 2000-675227 A 20000929
WO 2001-US30750 W 20010928

AB The invention provides isolated nucleic acid and amino acid sequences for human kinesin KinI-3. The invention also relates to antibodies to KinI-3. The invention also relates to methods of screening for KinI-3 modulators using biol. active KinI-3, and kits

for screening for KinI-3 modulators. The invention shows that kinesin KinI-3 is unregulated in lung, colon and breast cancers.

L28 ANSWER 5 OF 10 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 3
ACCESSION NUMBER: 2002:351338 HCAPLUS
DOCUMENT NUMBER: 137:74991
TITLE: A mechanism for microtubule depolymerization by KinI kinesins
AUTHOR(S): Moores, Carolyn A.; Yu, Ming; **Guo, Jun ; Beraud, Christophe; Sakowicz, Roman**; Milligan, Ronald A.
CORPORATE SOURCE: Department of Cell Biology, CB227, The Scripps Research Institute, La Jolla, CA, 92037, USA
SOURCE: Molecular Cell (2002), 9(4), 903-909
CODEN: MOCEFL; ISSN: 1097-2765
PUBLISHER: Cell Press
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Whereas most kinesins motor along microtubules, KinI kinesins are microtubule depolymg. machines. Surprisingly, we found that a KinI fragment consisting of only the motor core is capable of ATP-dependent depolymn. The motor binds along microtubules in all nucleotide states, but in the presence of AMPPNP, microtubule depolymn. also occurs. Structural characterization of the products of AMPPNP-induced destabilization revealed a snapshot of the disassembly machine in action as it precisely deformed a tubulin dimer. While conventional kinesins use the energy of ATP binding to execute a "power-stroke," KinIs use it to bend the underlying protofilament. Thus, the relatively small class-specific differences within the KinI motor core modulate a fundamentally conserved mode of interaction with microtubules to produce a unique depolymg. activity.

REFERENCE COUNT: 29 THERE ARE 29 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L28 ANSWER 6 OF 10 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN
ACCESSION NUMBER: 2003:156397 BIOSIS
DOCUMENT NUMBER: PREV200300156397
TITLE: Interaction of a KinI motor core with microtubules.
AUTHOR(S): Moores, C. A. (1); Hekmat-Nejad, M.; Yu, M.; **Sakowicz, R.**; Milligan, R. A. (1)
CORPORATE SOURCE: (1) Dept. of Cell Biology, Scripps Research Institute, La Jolla, CA, USA USA
SOURCE: Molecular Biology of the Cell, (Nov. 2002, 2002) Vol. 13, No. Supplement, pp. 280a. print.
Meeting Info.: 42nd Annual Meeting of the American Society for Cell Biology San Francisco, CA, USA December 14-18, 2002 American Society for Cell Biology
. ISSN: 1059-1524.
DOCUMENT TYPE: Conference
LANGUAGE: English

L28 ANSWER 7 OF 10 SCISEARCH COPYRIGHT 2003 THOMSON ISI on STN
ACCESSION NUMBER: 2002:563016 SCISEARCH
THE GENUINE ARTICLE: 570RY

TITLE: Potential long-term toxicity of repeated orally administered doses of artemether in rats.
AUTHOR: Xiao S H; Yang Y Q; You Q Q; Utzinger J; Guo H F; Jiao P Y; Mei J Y; **Guo J**; Bergquist R; Tanner M (Reprint)
CORPORATE SOURCE: Swiss Trop Inst, POB, CH-4002 Basel, Switzerland (Reprint); Swiss Trop Inst, CH-4002 Basel, Switzerland; Chinese Acad Prevent Med, Inst Parasit Dis, Shanghai 200025, Peoples R China; Princeton Univ, Off Populat Res, Princeton, NJ 08544 USA; WHO, WHO Special Programme Res & Training Trop Dis, World Bank, UNDP, CH-1211 Geneva 27, Switzerland
COUNTRY OF AUTHOR: Switzerland; Peoples R China; USA
SOURCE: AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE, (JAN 2002) Vol. 66, No. 1, pp. 30-34.
Publisher: AMER SOC TROP MED & HYGIENE, 8000 WESTPARK DR, STE 130, MCLEAN, VA 22101 USA.
ISSN: 0002-9637.
DOCUMENT TYPE: Article; Journal
LANGUAGE: English
REFERENCE COUNT: 29
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

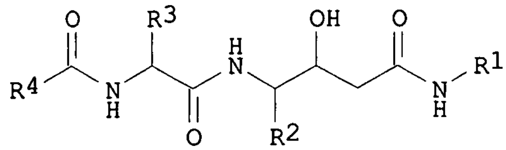
AB Artemether, an efficacious antimalarial drug, effectively prevents patent schistosome infections and morbidity, as established in laboratory models and in clinical trials. In view of concern about the potential long-term toxicity, rats were treated orally with 80 mg/kg artemether once every 2 weeks for 5 months. After the final treatment, routine blood test results were normal except for reversible reductions of reticulocyte counts and reversible increases in hemoglobin levels. Liver and kidney function and histopathological examination showed no differences between treated and untreated rats. Administration of 400 mg/kg artemether resulted in transient focal vesicle degeneration of the liver or slight damage to the proprius layer lamina of intestinal villi. No damage to the central nervous system tissues, including cerebrum, cerebellum, midbrain, thalamus, pons, medulla oblongata, and spinal cord, was seen at either concentration. There were no alterations in electrocardiograms during the 6-month treatment period. We conclude that 80 mg/kg artemether administered once every 2 weeks is safe, and doses of 400 mg/kg do not result in evidence of neurotoxicology.

L28 ANSWER 8 OF 10 SCISEARCH
ACCESSION NUMBER: 2000:878709 SCISEARCH
THE GENUINE ARTICLE: 374GN
TITLE: A statistical-based approach to assessing the fidelity of combinatorial libraries encoded with electrophoric molecular tags. Development and application of tag decode-assisted single bead LC/MS analysis
AUTHOR: Dolle R E (Reprint); **Guo J**; OBrien L; Jin Y; Piznik M; Bowman K J; Li W N; Egan W J; Cavallaro C L; Roughton A L; Zhao Q; Reader J C; Orlowski M; JacobSamuel B; Carroll C D
CORPORATE SOURCE: PHARMACOPEIA INC, DEPT CHEM, POB 5350, PRINCETON, NJ 08543 (Reprint); PHARMACOPEIA INC, DEPT ANALYT CHEM, PRINCETON, NJ 08543; PHARMACOPEIA INC, DEPT BIOL, PRINCETON, NJ 08543; PHARMACOPEIA INC, CTR INFORMAT & DRUG DISCOVERY, PRINCETON, NJ 08543

COUNTRY OF AUTHOR: USA
SOURCE: JOURNAL OF COMBINATORIAL CHEMISTRY, (NOV-DEC 2000) Vol. 2, No. 6, pp. 716-731.
Publisher: AMER CHEMICAL SOC, 1155 16TH ST, NW, WASHINGTON, DC 20036.
ISSN: 1520-4766.
DOCUMENT TYPE: Article; Journal
FILE SEGMENT: PHYS; LIFE
LANGUAGE: English
REFERENCE COUNT: 78
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS

AB A statistical sampling protocol is described to assess the fidelity of libraries encoded with molecular tags. The methodology, termed library QA, is based on the combined application of tag decode analysis and single bead LC/MS. The physical existence of library compounds eluted from beads is established by comparing the molecular weight predicted by tag decode with empirical measurement. The goal of sampling is to provide information on overall library fidelity and an indication of the performance of individual library synthons. The minimal sampling size n for library QA is 10x the largest synthon set. Data are reported as proportion (p) +/- lower and upper boundary (lb-ub) computed at the 95% confidence level (alpha = 0.05). As a practical demonstration, library QA was performed on a 25 200-member library of statine amides (size = 40 x 63 x 10). Sampling was conducted three times at n similar to 630 beads per run for a total of 1902 beads. The overall proportions found for the three runs were consistent with one another: p = 84.4%, lb-ub = 81.5-87.2%; p = 83.1%, lb-ub = 80.2-85.95; and p = 84.5%, lb-ub = 81.8-87.3%, suggesting the true value of p is close to 84% compound confirmation. The performance pi of individual synthons was also computed. Corroboration of QA data with biological screening results obtained from assaying the library against cathepsin D and plasmepsin II is discussed.

L28 ANSWER 9 OF 10 HCAPLUS DUPLICATE 4
ACCESSION NUMBER: 1998:630381 HCAPLUS
DOCUMENT NUMBER: 129:331033
TITLE: Identification of potent inhibitors of **Plasmodium falciparum** plasmepsin II from an encoded statine combinatorial library
AUTHOR(S): Carroll, Carolyn DiIanni; Patel, Hitesh; Johnson, Theodore O.; Guo, Tao; Orlowski, Marc; He, Zhen-Min; Cavallaro, Cullen L.; **Guo, Joan**; Oksman, Anna; Gluzman, Ilya Y.; Connelly, James; Chelsky, Daniel; Goldberg, Daniel E.; Dolle, Roland E.
CORPORATE SOURCE: Department Biology, Pharmacopeia, Inc., Princeton, NJ, 08540, USA
SOURCE: Bioorganic & Medicinal Chemistry Letters (1998), 8(17), 2315-2320
CODEN: BMCLE8; ISSN: 0960-894X
PUBLISHER: Elsevier Science Ltd.
DOCUMENT TYPE: Journal
LANGUAGE: English
GI



I

AB An encoded 13,020-member combinatorial library I [R1 = Me, Bu, MeO(CH2)3, Ph(CH2)4, PhCH2, 2-pyridylmethyl, 3-pyridylmethyl; R2 = CH2Ph, Me, CH2CHMe2; R3 = side chain from 31 natural and unnatural L- and D-amino acids; R4 = Ph, 2-naphthylmethyl, 2,4-(MeO)2C6H3, cyclopropyl, 2-carboxy-4-methylcyclohexyl, 4-(4-HOC6H4)C6H4, 5-phenyl-2-pyrrolylethyl, Me3C, 4-methylcyclohexyl, 2-naphthyloxymethyl, Et2NCH2CH2, 4-(Me2NCH2)C6H4, n-heptyl, 3-tetrahydrofuryl, 4-[(HO2C)2CHO]C6H4, HOCHMeCH2, 2,4-Cl2C6H4O(CH2)3, PhCH2] was synthesized containing a statine core. Evaluation of this library with plasmepsin II, an aspartyl protease required for Hb metabolism in the malaria parasite, led to the identification of potent and selective inhibitors as well as novel structure-activity relationships.

REFERENCE COUNT: 18 THERE ARE 18 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L28 ANSWER 10 OF 10 MEDLINE on STN

ACCESSION NUMBER: 93338845 MEDLINE

DOCUMENT NUMBER: 93338845 PubMed ID: 1307277

TITLE: Effect of sodium artesunate on **Plasmodium** yoelii analysed by flow cytometry.

AUTHOR: Han W S; Zhou W Q; Wang B X; Li Z M; Zuo L F; **Guo J W**

CORPORATE SOURCE: Department of Parasitology, Hebei Medical College, Shijiazhuang.

SOURCE: CHUNG-KUO CHI SHENG CHUNG HSUEH YU CHI SHENG CHUNG PING TSA CHIH CHINESE JOURNAL OF PARASITOLOGY AND PARASITIC DISEASES, (1992) 10 (3) 204-7. Journal code: 8709992. ISSN: 1000-7423.

PUB. COUNTRY: China

DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)

LANGUAGE: Chinese

FILE SEGMENT: Priority Journals

ENTRY MONTH: 199309

ENTRY DATE: Entered STN: 19930917
Last Updated on STN: 19930917
Entered Medline: 19930902

AB The effect of sodium artesunate on **Plasmodium** yoelii-infected mouse erythrocytes was analysed by flow cytometry. The results showed that malarial DNA content in experimental group was obviously decreased 2-5 hours after the drug was administered, fluorescence distribution of malarial DNA almost disappeared within 24 hours after the administration. Thus we deem that sodium artesunate can inhibit the DNA synthesis in P. yoelii.

FILE 'HOME' ENTERED AT 14:43:32 ON 07 OCT 2003

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         October  2, 2003, 16:13:56 ; Search time 130 Seconds
                                              (without alignments)
                                              2556.706 Million cell updates/sec

Title:          US-10-006-780-2
Perfect score:  6842
Sequence:       1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_23:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6842	100.0	1351	5	Q8I4Y0	Q8i4y0 plasmodium
2	903.5	13.2	2577	5	Q8IAM1	Q8iam1 plasmodium
3	876.5	12.8	3207	5	Q8IHQ0	Q8ihq0 plasmodium
4	857	12.5	3248	5	Q8ICP9	Q8icp9 plasmodium
5	854.5	12.5	2235	5	Q8IB09	Q8ib09 plasmodium
6	851	12.4	5561	5	Q8ILC9	Q8ilc9 plasmodium
7	850.5	12.4	1422	5	Q8MMY3	Q8mmy3 dictyosteli
8	847.5	12.4	1681	5	Q8I548	Q8i548 plasmodium
9	839	12.3	3178	5	Q8I2P8	Q8i2p8 plasmodium
10	838	12.2	828	5	Q8MN35	Q8mn35 dictyosteli
11	838	12.2	3256	5	Q8IDV7	Q8idv7 plasmodium
12	830.5	12.1	3069	5	Q8I350	Q8i350 plasmodium
13	830	12.1	2874	5	Q8ICX0	Q8icx0 plasmodium
14	827.5	12.1	2432	5	Q8I5I9	Q8i5i9 plasmodium
15	826.5	12.1	1824	5	Q8I5A3	Q8i5a3 plasmodium
16	826.5	12.1	2672	5	Q8IEE4	Q8iee4 plasmodium
17	824	12.0	901	5	Q8T806	Q8t806 dictyosteli
18	819.5	12.0	1483	5	Q8IIQ7	Q8iiq7 plasmodium
19	819.5	12.0	4466	5	Q8ILS4	Q8ils4 plasmodium
20	817.5	11.9	2275	5	Q8IHV8	Q8ihv8 plasmodium
21	816	11.9	2041	5	Q8IBK6	Q8ibk6 plasmodium
22	815.5	11.9	1774	5	Q8IAU8	Q8iau8 plasmodium
23	813	11.9	3848	5	Q8IDR0	Q8idr0 plasmodium
24	810.5	11.8	2472	5	Q8T2M5	Q8t2m5 dictyosteli
25	810	11.8	1200	5	O97277	O97277 plasmodium
26	808	11.8	2612	5	Q8I5X5	Q8i5x5 plasmodium
27	807	11.8	1918	5	Q8SSW3	Q8ssw3 dictyosteli
28	804	11.8	2206	5	Q8IBU8	Q8ibu8 plasmodium
29	804	11.8	8591	5	Q8IB94	Q8ib94 plasmodium
30	803	11.7	1785	5	Q8I5Z0	Q8i5z0 plasmodium
31	802	11.7	2763	5	Q8II48	Q8ii48 plasmodium
32	801	11.7	1770	5	Q8I3A3	Q8i3a3 plasmodium
33	800	11.7	2940	5	Q8IHP9	Q8ihp9 plasmodium
34	799.5	11.7	3569	5	Q8IBV8	Q8ibv8 plasmodium
35	798	11.7	1605	5	Q8I4X5	Q8i4x5 plasmodium
36	797.5	11.7	1922	5	Q8IL84	Q8il84 plasmodium
37	796	11.6	800	10	Q94GW1	Q94gw1 oryza sativ
38	795.5	11.6	706	10	Q9LW81	Q9lw81 arabidopsis
39	795.5	11.6	1192	5	Q8IEI5	Q8iei5 plasmodium
40	795	11.6	1437	5	Q8ICZ4	Q8icz4 plasmodium
41	794	11.6	684	10	Q940Y8	Q940y8 arabidopsis
42	793.5	11.6	2849	5	Q8IHY4	Q8ihy4 plasmodium
43	791	11.6	3334	5	Q8IIS9	Q8iis9 plasmodium
44	790.5	11.6	703	10	Q9AVP4	Q9avp4 nicotiana t
45	789.5	11.5	799	10	Q9LUS1	Q9lus1 arabidopsis

ALIGNMENTS

```
RESULT 1
Q8I4Y0
ID   Q8I4Y0      PRELIMINARY;      PRT;  1351 AA.
AC   Q8I4Y0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Kinesin-like protein, putative.
GN   PFL2165W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014851; AAN36517.1; -.
SQ   SEQUENCE   1351 AA;  156007 MW;  8B69594BFD73C788 CRC64;

  Query Match             100.0%;  Score 6842;  DB 5;  Length 1351;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-268;
  Matches 1288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 123
```

```
Qy      61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     124 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 183

Qy     121 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     184 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 243

Qy     181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     244 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 303

Qy     241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     304 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 363

Qy     301 GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGS 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGS 423

Qy     361 EMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKY 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     424 EMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKY 483

Qy     421 SSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSS 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     484 SSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSS 543

Qy     481 SMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 SMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 603

Qy     541 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     604 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 663

Qy     601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     664 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 723

Qy     661 NGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNK 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     724 NGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNK 783

Qy     721 NNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNS 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     784 NNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNS 843

Qy     781 HNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIH 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     844 HNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIH 903

Qy     841 GCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNE 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     904 GCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNE 963

Qy     901 MKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKN 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     964 MKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKN 1023

Qy     961 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLD 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1024 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLD 1083

Qy    1021 DQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIK 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1084 DQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIK 1143

Qy    1081 NFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYIN 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1144 NFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYIN 1203

Qy    1141 EDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1204 EDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN 1263

Qy    1201 NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSN 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1264 NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSN 1323

Qy    1261 NGDIILLNKKLVQDNIKNSMDHNNIHKK 1288
           ||||||||||||||||||||||||||||
Db    1324 NGDIILLNKKLVQDNIKNSMDHNNIHKK 1351
```

```
RESULT 2
Q8IAM1
ID   Q8IAM1      PRELIMINARY;      PRT;  2577 AA.
AC   Q8IAM1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   MAL8P1.153.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA   Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844507; CAD51342.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2577 AA;  299292 MW;  025BBC240587F4D7 CRC64;

  Query Match             13.2%;  Score 903.5;  DB 5;  Length 2577;
  Best Local Similarity   23.1%;  Pred. No. 1.1e-28;
  Matches  430;  Conservative  228;  Mismatches  482;  Indels  721;  Gaps   89;

Qy      19 KKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIER----HEFIVDKVFDDTV--DNFTVY- 71
           |||| |    : :           |::  |: |:    :    :: ::| :  |||: |
Db     111 KKKKKSRYCDIND-----------KLNFVKFNEKRKCTNRSYNNRYYNDYLKNDNFSKYL 159

Qy      72 --ENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAAGDIFT 129
             :|  |                    :   |:  :  |:   :       :|:  ||
Db     160 SSKNNKK------------------NSNLKKSSNLFISEEQPEQGQEDFLEYSEQDIIK 200

Qy     130 FLNIY--------------------DKDNTKGIFISFYEIYCGKLY-DLLQKRKMVAAL 167
       .    :  :                     |: |       : ::|  | | |   |||
Db     201 LIYYFMENFIFKKKKENRQMNGKQLGDRTN------EYVDMYTKKSYSDKNVKRK----- 249

Qy     168 ENGKKEVVVKDLKILRVLTK-----------------------EELILKMID--GVLLR 201
           : ||  : ||::  : |: |                        :::|:| ||  |
Db     250 KEGKINIKVKNISQMNVINKNVELAILGRDSLSLSLPEEVKGENDQVIIKEIDEKGKRDD 309

Qy     202 KIG--VNSQNDESSRSHAILNIDLKD----INKNTSLGKIAFIDLAGSERGADTVSQNKQ 255
           |||  :| :| : :  : :|  : :|    |||   | |         |   : : :
Db     310 KIGTVLNDENVKRNLLNGLLVCNEEDNEEVINKEVLLLK---------ENNINDIMKKDV 360

Qy     256 TQTDGAN------------INRSLLALKECIRAMDSDKNHIPFR---------------- 287
            :|: :|            || : :  ||   :  ||  ||
Db     361 YKTNMSNDNKLNMENIENDINLTSICYKEKNSSEPSDVTHINIEYEGKGIPNNGDTIDEN 420

Qy     288 ---DSELTKVLRD----IFVGKSKSIM---IANISPTIS-------CCEQT--------- 321
              | :  |:::|    : :| :  |:     |::  ||       | |::
Db     421 IIDDEKENKIIKDEKEYMEIGNNDDIINDCKKNMNKEISMKLGMKGCVEESNNISEAVKE 480

Qy     322 ----------------------LNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK- 358
                                 :| : |   :|||   : | |: :    :| :| : :
Db     481 EEKLKILQIIDNYLQKNEMFFNMNDIFYPFTIKNF---NLCKNKNEKNKMKRNTITERRY 537

Qy     359 -GSEMNASSIENVVIKSNHLLSNNNN----NKINRGKINDK-----IERNN------ILK 402
            |:: |  :| |    : ::  ||||    ||:     | |      | ||       :|
Db     538 LGNKNNNINIYN---NNINMYYNNNNISSLNKVQCNMSNSKGVYISEESNNNCPSKEYIK 594
```

```
Qy    403 NK-SFDKPREG-----------------------FTSTFGKYSSLNDID----------K 428
Db    595 NKIYFSKLRMGLNMLLKHEKKSKRGRRTGGASKNYKKELGEQKKKDGMDCDMLVRDSSKR 654

Qy    429 IKKNKKKGLI---------NYKSTL-YNDNTINKKHNNNNNNNNDNDNNNDNNNNNNND 478
Db    655 VNKSRNKNKLPDNNRNNNNNMKDNLDINMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 714

Qy    479 ------------------------------------------------------------ 478
Db    715 YYVYGGGATNIGNGKNKFANCNLPYDNPKEYHNVINKDKNLIRRNILNVVDTSNMSMYNS 774

Qy    479 ----------------SSSMVNNMI---NHMINNNINNNINVNNNNNNNNNNNSHNNHL 519
Db    775 HNGDNFNILNKNEPSFSSTNTNNVILKSNHNNNNNNNNNNNNNNNNNNNNNNNNSYNNMI 834

Qy    520 P-----QPNYAFTDTSDFSSLDDMNC-----HLNNNDKSIFLHKKNLRDNI--------- 560
Db    835 TFVENNEPN----DTGNF--YNPKSCVPYYDKTNNNIKNVDDNYNQKNYNVVSFNNSQVH 888

Qy    561 -------KLKNRSS--CDNIMNKKKNNLHLARHSVGSKLTMF------------------ 593
Db    889 PPYNHEEEYKNNSSKVCNTYTNNYMHNINNVNYA-GSNMNACLNENSVKNDEDCKINYIN 947

Qy    594 --SYDPQKNKDNTFF-----KSNINKMEDNTP-KDILYESRNVSNMNGNV---------L 636
Db    948 NNNMEENKNIDNVIIRMYNKKSHTQRM--NVPTKDLKNE---VYNKTGHVKTCNVIKSKT 1002

Qy    637 LGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSN----MNSINN-SNMNSN 691
Db   1003 ISLN-NIVLDLSPNEQMTNGMYENEDMYNVRFKSDENYYNTMNMIGYMDKINNMNNLNVT 1061

Qy    692 SIYKSNYNSNQSISDVQIRY---VNEMDTSNK------NNDNIFFDAISCDNNMYPNITN 742
Db   1062 KPTRKNNKSNKNAGKDKTKSRDGKNNNDSSNSPQIFEDNHDEYKDD----DDN---NNNN 1114

Qy    743 NNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNK 802
Db   1115 NNNNNNNNNNNNNNNINVNNNGNNNGNKENHHGHKENGDN--KKDNDRN--NKENCGHN- 1169

Qy    803 NNQDGNVN--YSMN-------FCHYNLND------KNYL------IDLNNKEQKDKN--- 838
Db   1170 NNQDNDYNNTYKDNEGNNMTSFYNNNINTNYLDERENYLCMEQNNIYLQNEKGKNNNDAF 1229

Qy    839 ----------IHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNN--------------NM 874
Db   1230 LGPEQNCIATIGSCNNNNNDNTSNNMKNKLIRNMNFNGNVNVNNYNITDNDKCLYNVKNI 1289

Qy    875 GNNNSPRMKYGLCGSHTSIDNMKN-NEMKNNEMKDN----EMKDNH-------------- 915
Db   1290 INNNEKR------GNLCYIKNVKNMNNLHNNNILGNNTFERFDDNNTVHMNSYDDQLYNP 1343

Qy    916 -------------------IKSNNNNSSSSSSNN--NIYN--NINDDDTFQNDYCHNDN 952
Db   1344 LNEEEEKKKKKRKKKKSPTIINNDGNINNDGNINNDGNINNDGNINNDDNINNDGNINN 1403

Qy    953 TFTIRRKNNTNINSNIYQ-NDDIIYTINSLN--------------------DYMS----- 986
Db   1404 G-NINNDDNINNNGNIIQETQDDFYLTNHINNNHMMHKQNMSQNIIYHNKVDYLNENDSN 1462

Qy    987 -NTLLHFKEKYTYPTLSTNEDIYN-KEMEGKHIRLDDQDKYDDN---DNNNVDN------ 1035
Db   1463 MNLLDEKKIGYQKKTSKKNEGIMNGKKGKRKYQRKDNKNVNDSNVFTGNNSVFNILPSLY 1522

Qy   1036 --------NNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEY 1087
Db   1523 TKVENEIDVNNNNI-NNINNNNI-NNINNNNINNINNNNINN------INNNNNNNNSNI 1574

Qy   1088 LSYFQKNVDTIINNCLNSLDISSMYDDT---------------------KEILNNILLS 1125
Db   1575 CTYYNYNTYEVIYNDIHNDDKKKKKSDSKRKDRNSKKKINIEQKIENNNNKESFDNNEMN 1634

Qy   1126 KYKAEKDNVIKKYIN---------------------EDIKNMSLEEIDKTAQSIYEKRK 1163
Db   1635 NYNHELERVTYDSINNNGNIHKGDYIKNNVYNNNYHLHDVNTNSVSSLNNKASSF----N 1690

Qy   1164 VLLTKLLLLFKK---------NVDTQIN---------------NET---SDLRKDLV--- 1193
Db   1691 TIFTNSRSTFPKSLSFNSDVTNAKDEINVIGLANSISTGNSTGNSTECVKDGNKDLIGYI 1750

Qy   1194 --------------------------MCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQ 1226
Db   1751 QKVDEMDKAGHMNKVDFINSGNNDYKNKLDIINNKMYDQFNNVNTGRYHSSNI------- 1803

Qy   1227 IWCESENLRL---LYQFLVVEYQNKSANSVLLNVSSNN-------------GDIILLNK 1269
Db   1804 --CSDKNYRLNNCINKSMDLKY---STNNIIINNMNNNIFPNECNNKYVCGSDDNMFLNK 1858

Qy   1270 K 1270
Db   1859 K 1859
```

RESULT 3
Q8IHQ0

```
ID   Q8IHQ0      PRELIMINARY;      PRT;  3207 AA.
AC   Q8IHQ0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF11_0479.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014843; AAN36059.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   3207 AA;  377920 MW;  71C32BF88BF06A59 CRC64;

  Query Match             12.8%;  Score 876.5;  DB 5;  Length 3207;
  Best Local Similarity   23.3%;  Pred. No. 1.7e-27;
  Matches  388;  Conservative  245;  Mismatches  520;  Indels  513;  Gaps   76;

Qy      1 MNSKIKVVVRKRPLSELEKKKKDSDIITVK---------NNCTLYIDEPRYKVDMTKYIE 51
Db      5 IQEKLQEKKKKKKKMSDRKEDKNDIILNKNEEEDNINNNNIILYKSFDDFKINYSYKTK 64

Qy     52 RHEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCV----------CSCFAYGQTGSGK 101
Db     65 NHLHENDKIKEE--DDHEIKRKLIKLINTNFYIDKCIHFKKFSKDELYKTFIYSNFLTKA 122

Qy    102 TYTMLGSQPYGQS---------DTPGIFQYAAGDIFTF-----LNIYDK-----DNTKGIF 143
Db    123 LILYPSLMPYVECIIEKIKKIKNENITFFPAIEQFNFSIEHAVSSYQTGTQTFNNHPNFY 182

Qy    144 ISFYEIYC----------GKLYDLLQKRKMVAALENGKKEVVVKDLKILRVLTKEELILK 193
Db    183 TNYYQSFIKNDNIPYINQTNIFD-----------NNIKNKYMLDD--------------K 217

Qy    194 MIDGVLLRKIGVNSQNDE------------SSRSHAILNIDLKDINKNTSLGKIAFIDLA 241
```

```
Db    218 FGSTSLYNNNNNNNNNNENNNDKYLNTYYASPRGNQIYNL-FQDINNNHHNNNI------ 270
Qy    242 GSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVG 301
Db    271 ------NSYSNNKYFNYNMLNKNN----LEETRSYPNNDYRNI-INPNNL---------- 309
Qy    302 KSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSE 361
Db    310 NSNNVHMNNIYPELN---NTYNNMNNTYHLNNVHSTSLQNNYNPNNNFDH---------- 356
Qy    362 MNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNI--------------------- 400
Db    357 -SMKNIKNETNNNNNNNNNNNNNNNNNNNNNNNNNNNNNIYSPELDTHLRVYKNINEYDII 415
Qy    401 -------LKNKSF-------------------------DKPREGFTSTF---------- 417
Db    416 ELKICTTLNNKYICRYIGIHSGMTLYNIHRCICICLGVKKEDTDNYLHVFILNNGNIYGS 475
Qy    418 GKYSSLNDIDKIKKNKK-----KGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNN 472
Db    476 GKKCSVSIIRKIQINTDRHITFKHIIKTPLYLYKSKEDKNKNNSNNNNNNNNSNNNNNSN 535
Qy    473 NNNNNDSSSMVNNMI--NHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTS 530
Db    536 NNNNSNSNSNSNNNTNSNSNSNSNSNNNTNSNNNSNSNSNNNNSNNN-------TTTNNS 588
Qy    531 DFSSLDDMNCHLNNNDK--------------SIFLHKKNLRDNI---------------- 560
Db    589 SSSNNSNNNNYYHNNYKNEKELNNSSSLEHSSIIMNNDNIHNNINNNHINNITDLNNMNV 648
Qy    561 ---KLKNRSSCDNIMNKKKNNLHLARHSVGSKLT------------------MFSYDPQK 599
Db    649 NQSNMKENNNIIDYMNNNNNNDNYSNNHLNNCINKLYTNNIYFTEDSQKRNPLQTYNTSK 708
Qy    600 NKD-----NTFFKS----NINKMEDNTPKDILYESRNVSNMNGNVLLGLNK---NTHHDI 647
Db    709 NTNNFLNVNNFTSSYNFPNINNMDSN-----IYNHTTCNNFNKNINNNINDISINKHNNI 763
Qy    648 --STKDENHNDNK-------------------INNGVINIINNSNVNSIN-----NSNMN 681
Db    764 FNNMNHLNHLDNHSYIQNNLYKNHMNVNTNILYNNPIMNNINNDQINNLSIPNNKNEDNN 823
Qy    682 SINNSNMN---SNSIY--------------------KSNYNSNQSIS---DVQIRYVNE 714
Db    824 EINHDDSNDDDSNSSHITLNKSDKNKNYFALNPKYQNHQNHNINNNIQNNLNEQIKEKND 883
Qy    715 MDTSN----KN--------------NDNIFFDAISCD-------------------NNMY 737
Db    884 QQNHNIKEIKNKELLNDTISSIEDTNDNSYSKYITSSDISQNNTLNSFQHNKEISVNFMY 943
Qy    738 PNITNNNNNNNNNNNNNNNNIDVE--NYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNI 795
Db    944 NNIILDNNNNINDDNNNNNNYFCIPCGYNTKE--------YKYNIYNTYNYPNNAN-HIY 994
Qy    796 QNINTNKNNQDGNVNYSMNFCHYNLNDKNYLI------------DLNNKEQK-------- 835
Db    995 NNMNISYNNSAYNNNYVTYNNFHNSYHNNYILHNNFHNPYNIYDNIQNTEQKKLYNIYQN 1054
Qy    836 DKNIHGCDNNI------IQNRNDFEKKKKTNFYNNN---NIVIVNNNMG----NNNSPRM 882
Db   1055 DERQNNSFNHINTDPHKVVNSNNF-LPINTFHYNNNLNHNILTESNNLNRKNENDNIPSS 1113
Qy    883 -----KYGLCG-----SHTSIDNMKNNEMKNNEM-----KDNEMKDNHIKSNNNN----- 922
Db   1114 YSQIHNHQICKKVEEYTYNSI-NQNTNNFNNNVMMLMNTSNNIPLDNNTYNSNKNKIIYK 1172
Qy    923 ---SSSSSSSNNNI-YNNINDD-DTFQN--DYCHNDNTFTIRRKNNTNINSNI----YQN 971
Db   1173 HIINDHINQKDNNVEYENLNNSCDNTQNKETFCNQDLI------NSSNINNNISSYTFQN 1226
Qy    972 DDIIYT------INSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEG--KHIRLDDQD 1023
Db   1227 NNDFYTKKKSMQYNHDNIYKINTT---SENVGSPHTNNKTSIYNHKKGGYEQHTEQNNEQ 1283
Qy   1024 KYDDNDNNNVDNNNKNNVDNNVDNN---NVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIK 1080
Db   1284 NNEQNSEQNIEQNIEQNIEQNVAQNVAQNVAQNVEQNVEQNVAQNVEQNVEQNVEQKAEQ 1343
Qy   1081 NFNNNEY----LSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIK 1136
Db   1344 NSNNESIKTNTVETFKRNKNQITNS--NNV-ISKQQHDTNNILNNININ----IKENINR 1396
Qy   1137 KYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNE---TSDLRKDLV 1193
Db   1397 HKINE-------FQWEKSNKIDIEKNNCLTTK----YDKDNDNENDNENDNTYNKNNDIV 1445
Qy   1194 MCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEY-------- 1245
Db   1446 ICNNHNNSSHVQKNYY---NMNESMINENNI--IITEGENLMNSTE----EYFTNELIKK 1496
Qy   1246 ----QNKSANSVLLNVSSNNGDIILLNKKLVQDNIKNSMDHNNIHK 1287
Db   1497 DSLEKNKSDTKFLIKL---NNEIKKEEEK--KDNI-NIFINNNIYE 1536
```

RESULT 4
Q8ICP9

```
ID   Q8ICP9      PRELIMINARY;      PRT;  3248 AA.
AC   Q8ICP9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   MAL6P1.307.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844505; CAD50391.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   3248 AA;  374892 MW;  F72C129A97E38466 CRC64;

  Query Match             12.5%;  Score 857;  DB 5;  Length 3248;
  Best Local Similarity   23.0%;  Pred. No. 1e-26;
  Matches  347;  Conservative  220;  Mismatches  413;  Indels  530;  Gaps   62;

Qy    202 KIGVNSQNDES-----SRSHAILN----------IDLKDINKNTSLGKIAFIDLAGSERG 246
Db    371 KVTIGSYNNEDNKFNMNKMHVIKNEMDGNANDNKIENEDVN----------IPICGSSKS 420
Qy    247 ADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSI 306
Db    421 PLNIQNN--------NI-----------------EINH-----------MRDNFL--SNDI 443
Qy    307 MIANISPTISCCEQTLNTLRYSSRVKNFKNKS----------TCINEEDDTNTERISILD 356
Db    444 M--------------NTIPYSPPTKKIKIEKRMMMKNIMHPNVIKNMDENNMD-MDKMS 487
Qy    357 SKGSEMNASSIENV------------------------------------VIKSNHLL 378
Db    488 ANNIDMNKMSVNNIDMNKMCATNVDINKMSVNNIDINKMCGNNIDINKMCGNNIDMDHMN 547
Qy    379 SNNNN----NKIN-----RGKINDKIERNNILKNKSFDKPR----EGFTSTFGKYS---- 421
Db    548 SNNNDLINMNKYNIHMNINNKIPNKFDANENIRIGKTDHPKYLNGKDYAYTRSCYNENNI 607
Qy    422 --SLNDIDKIKKNK----KKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNN 475
Db    608 EYNENENNDIKGNKNIIRKNNIMNFNTNNNNNNNNDNDSSNNNNNSNNNNNSNNSNNNINN 667
```

```
Qy        476 NNDSSSMVNNMINHMINNNINNNINVNNNNNNNN-----NNNNSHNNHLPQPNYA----- 525
Db        668 NNSNNSNNNNNNNNNNNNNNNNNNNNNKNNNNNSNNNNNSYNNNNSNNNNNAQITYGENIYS 727

Qy        526 --------FTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNK--- 574
Db        728 TSNYQNNNFYEKNNFVEIKRININENFCDEK----NKELKKNMVPKNQDTLNEVTNKILI 783

Qy        575 --------KKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESR 626
Db        784 NKNAELTNQKNLKKKMRANINACLS--GSNNNINNNNIIFYNEFIK-DDNYYNNICLNNE 840

Qy        627 NV-SNMNGNVLLGLNKNTH----HDIST--KDENHNDNKI--NNGVINIINNSNVNSINN 677
Db        841 ELNNNNNNNILFNDNKNCNYKNLHLNSTINFNRNNNNNNIDNNNNNININNNININNIKN 900

Qy        678 SNMNSINNSNMN--------------------------------------SNSIYK-- 695
Db        901 NNDYYIKYNNLNEFKEEEEKEEKEKEKEKKKIEKKDNEKEKEKLPNEQEEEQQREIIKKG 960

Qy        696 --SNYNSNQSISDVQIRYVNEMDTS------NKNNDNIF---------FDAISCDNNM-- 736
Db        961 KTQNINSNKKIT----KCINSSNSSSYLMNYDNNNNNYIKENTKNLAEFQNINNINNLTS 1016

Qy        737 --YPNITN-------------NNNNNNNNNNNNNNIDVENYN----NRDGTNNSMKLY 776
Db       1017 NNCPHVLNVSSLNVKNMENQMARNNSSFNNINNNHLNNDMHNFNQHLHNLKHINNPINL- 1075

Qy        777 AYNSHNLFQPDN-----------------------NKNTSNIQNINTNKNNQDGNVNYS 812
Db       1076 --STHGSLQKNNVQMRMNTKKIVPEENIEMESFMNINKNSSNINFIEHNMMN--NNIN-- 1129

Qy        813 MNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNN 872
Db       1130 ---AYYNNNNNNMI--NNGTNETPKRTHRASLNIMQNINIPQKNNFTLNSNKNDVV---- 1180

Qy        873 NMGNNNSPRMKYGLCGSHTSIDNMKN---------NEMKNNEMKD-------------- 908
Db       1181 -MNNINSYHVQ------HVNYEDMENKNAYPKNFINNLENHQMTTTIDGNKKRVSEHSGD 1233

Qy        909 -NEMKDNHIKS--NNNNSSSSSSSNNNIYNNIN----------------DDDTFQNDY-- 947
Db       1234 LNRMLNNNLNTPPNNINNTNNINNINNINNNVNIVKGKTVQMNSPLKNMTNSLHANNYNI 1293

Qy        948 ------------CHNDNTFTIRRKNNTN--------INSNIYQNDDIIYT---------- 977
Db       1294 NKMMVYPPQNIITHGNTTTNITTTNNNNNNNNNNSGYNNNVYPYDNKVDNSNNNPTNNDF 1353

Qy        978 INSLNDYMSNTLLHFKEKYTYPTLSTNEDIY-------------NKEMEGKHIRLDDQD 1023
Db       1354 INHLNNDTQNRKKYLASSHAANMQNSGNHVYINNISINNNVNMMKNNMNNNNNMISKNS 1413

Qy       1024 KYDDNDNNNVD---------------NNNKNNVDNNVDNNNVDN------------NVDN 1056
Db       1414 LANINNNNNMNNMNNKMNHMNNMNNMNNNMNNLNNSINNNNMNNINVGNNVLVGYKNVSN 1473

Qy       1057 NDKN-NVDNNNV-------------DNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIIN-N 1101
Db       1474 NNRNINNDNNNVILNPNYDMISRNENTQNNNLNMANVQNVYSNNNTGGFLVNKDYVVNKN 1533

Qy       1102 CLNSLDISS-------------MYDDTKEILNNILLSKYKA----EKDNVIKKYINEDIK 1144
Db       1534 MINENKVSNLETTHPYNHIDNKMMTDFNNNNNNSNINNNKSGANINYNNMYNQNIHNRMN 1593

Qy       1145 NMSLEEIDKTAQSI------------------YEKRKVLLTKLLLLFKKNVD--------- 1178
Db       1594 NTNTIVQNNVRSSILLNNKNNSNSSNNNNLPYDKENVTNIKYDLLNNKNVHNNIPFIGTK 1653

Qy       1179 -------TQINNE------TSDLRKDLVMCHICNNNP-----DDQFHFYAYSRLEKDIIN 1220
Db       1654 INLNNSLVQLNNNMNKNILSEDINKNGDILNSFNGNALPINYDRNEYASSYSEKNTNINE 1713

Qy       1221 LIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKLVQ----DNI 1276
Db       1714 TYMNSMINSKNNSIHLNQMF--------TKNTITSNKNSSNEKQIMNRKKLTKTQNINNL 1765

Qy       1277 KNSMDHNNIH 1286
Db       1766 KATDINSNIN 1775
```

RESULT 5

```
Q8IB09
ID   Q8IB09      PRELIMINARY;      PRT;  2235 AA.
AC   Q8IB09;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Asparagine-rich antigen.
GN   PF08_0060.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA   Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844507; CAD51201.1; -.
SQ   SEQUENCE   2235 AA;  254753 MW;  F223CA8261E0C7F3 CRC64;

  Query Match             12.5%;  Score 854.5;  DB 5;  Length 2235;
  Best Local Similarity   25.9%;  Pred. No. 9.3e-27;
  Matches  332;  Conservative  183;  Mismatches  438;  Indels  329;  Gaps   55;

Qy        206 NSQNDESSRSHAILNIDLKDINKNT----SLGKIAFIDLAGSERGADTVSQN-------- 253
Db        963 NNNNNNNSGSIVNLGNNNMDISMNSKTLYNLSKLSCATNSTTDSSKMSLKRNIWYNNRSN 1022

Qy        254 -KQTQTDGA-NINRS-----LLALKECIRAMDSDKNHIPFRD----------SELTKVLR 296
Db       1023 SVHSALNGALNSNMSNNYSDNMLLRSSIKNGPKYSVHYPSSSNNQDTTTMSISNKKSVVR 1082

Qy        297 DIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILD 356
Db       1083 NLMLNTQDQNFVLNNMYNNNSSNNNTNNSNNTNNTNNSNNSSSSNNNNTSSATYNNYMS 1142

Qy        357 SKG---------SEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNN----ILKN 403
Db       1143 SYGYVNMPTNNNINMNENNVEFNNTKKNSVTDNLNNN----GNVKDLLTNGNMNNAVYYN 1198

Qy        404 KSFDKPREGFTSTFGKYSSLNDIDKIKKN-----KKKGLINYK---STLY-NDNTINKKH 454
Db       1199 KSISSNNNGKDLRNNFYNNSN-IDLNKNNCYSIDNKQQLANSKFVNNNIYPNYNNMNDMT 1257

Qy        455 NNNNNNNNDNNDNNNDNNNNNNNDSSMVNNMINHMINNNINNNINVNNNNNNNNNNNS 514
Db       1258 NNNNNNNNNNNNNNNNNNNNNNNN-------------NNNNNNNNNNNNNNSNNNNDYV 1304

Qy        515 HNNHLPQPNYAFTD---TSDFSSLDDMNCHLNNNDKSIFLHKKN--LRDNIKLKNRSSCD 569
Db       1305 HNNINVSTSLGCATGSVMSQGSSTNNMNYNLRNYESKNFNSYQNGIISDNSNSNNNSNNN 1364

Qy        570 NIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVS 629
Db       1365 NNNNNNNNN----TNSSSSNGLIYHY------------PNYDKIKTNAPKGILNEYKLVK 1408

Qy        630 NMNGNVLLGLNK----NTHHDISTKDENHNDNKI--NNGVINIINNS---NVNSINN-SN 679
Db       1409 NVRSNSNSLNNKKGSNNTYYNTKT-SQNAYQNTILKNSPLVSLSNNSGMMNYNMINDMKD 1467

Qy        680 MNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPN 739
Db       1468 ENNMNAGNMYTGSISSSNNN-------------NSNNNSNNNSN---STMTSNGLLYNN 1511

Qy        740 ITNNNNNNNNNNNNNNN----------------IDVENYNNRDGTNNSM----KLYAYN 779
```

```
                : ||||  :  |: |||                      |  ||||  :  || :    |::  |
Db       1512 MNIINNNNVDELNDINNNKVGIPPHHLLDNTSLIKDEMNYNNEE--NNMVSQHHKVFIGN 1569

Qy        780 SHNLFQPDNNKN----TSNIQNINTNKNN-----QDGNVNYSMNFCHYNLNDKNYLIDLN 830
              |     ||::|      ||    |: | ||      :| | |  :|    : | |: |     |
Db       1570 GHENMLKDNSENLLKHMSNDTKIDENNNNTHIDDKDYNKN-NMQLDYSNNNNNN-----N 1623

Qy        831 NKEQKDKNIHGCDNNIIQNRNDFEKKK----KTNF------------------------ 861
              ||     :|   |  |          || |:     ||
Db       1624 NKMDHSRNGSGSD--------DFMKRKYKKMNTNFDSNNEHLNDEFKDIDIVDAQNDDTD 1675

Qy        862 --------YNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKD 913
                      |   |||:|  :  |          |  |     :  ||       :  ::||     :
Db       1676 KLRDINGKYEENNIIIDDINSKERRSINEYNNKCSSTAAYSEMK----QYVDVKDTNNHN 1731

Qy        914 NHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDD 973
              |::  ||   |  :   :::||| ||  |:::|: ||       ||   |    :|
Db       1732 NNMLSNYLNMTDENNNNNN--NNNNNNNTYLND---DNVKTTEKKRK------------ 1773

Qy        974 IIYTINSLN-DYMSNTLLHF----KEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDN 1028
               : |  |    | | |:| |  |          : :     : : :   :|: :|   |:::     :|||
Db       1774 FLNTYNDPNVDDMNNNLSHHVGVDYDNHVVDSFNNSSYMYSSQMY-KYMK---SNKYDTC 1829

Qy       1029 DNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNV-DNDDDDVDFHNIKNFNNNEY 1087
              | |::| |   | :||  :||| :|| :|:  ||::  ||| ||    :  ||   :      ||:
Db       1830 MNTNMENINGTNNNNNNNNNNNNNNNNNNNNNINRNNVSDNSTNIVDNDSYDKTNNSSN 1889

Qy       1088 LSYFQKNVDTIINN-----CLNSLDISSMYDDTKEILN---NILLSKYKAEKDNVIKKYI 1139
               | ::  ||   ::||          ||    |                :||      |:     |    :   :|
Db       1890 GSMYKTNVQNMVNNNNGAATLNKCD---------SLLNHDMNVSKLNYNSNNNN------ 1934

Qy       1140 NED-----IKNMSLEEIDKTAQSI-------YEKRKVLLTKLLLLFKKNVDTQINNETS- 1186
              | :       :    |::         || |:           ||      |             :||   ||    :
Db       1935 NSNNNSCGVAGGSMDSHYNTANSLMNYNKNHYEDGNDL---------NNMDTNNNNNNNN 1985

Qy       1187 ---------------------------DLRKDLVM--------CHICNNNP-----DDQF 1206
                                        |  :| : :|        |     :|||        :|:
Db       1986 NGSNSNNSNSNSNSSNNNNNKNKNLMGRDSLQDKIMYNPMNDYNCIYPSNNPYGIQNNDKG 2045

Qy       1207 HFYAYSRLEKDIINLIMLRQIWCESE----NLRLLYQFLVVEYQNKSANSVLLNVSSNNG 1262
               : |            :|:|          :|     |   ::    ::    | |  :          | ::||
Db       2046 YMYPNYNTALNIMN---------PNENGDNNNSMVGASVMDNYNNNNLYMNNNNNNNNNN 2096

Qy       1263 DIILLNKKLVQDNIKNSMDHNN 1284
               :  |  |          ||:||    :||
Db       2097 NNYLNNYMNPTDNMKNMSKNNN 2118
```

RESULT 6
Q8ILC9

```
ID   Q8ILC9     PRELIMINARY;     PRT;  5561 AA.
AC   Q8ILC9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF14_0315.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014821; AAN36928.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   5561 AA;  658524 MW;  98BF06CAC3C2B52B CRC64;

  Query Match             12.4%;  Score 851;  DB 5;  Length 5561;
  Best Local Similarity   23.3%;  Pred. No. 3e-26;
  Matches  385;  Conservative  212;  Mismatches  484;  Indels  568;  Gaps   73;

Qy         19 KKKKDSDIITVKNNCTLYIDEPRYKVDMTKYI---ERHEFIVDKVFDDTVD--------- 66
              |:|   :  ||     |   ::::  |         |       |:    :|  ::     ::
Db        521 KQKNTYNYITKSYLCVSFLNQLYYSKTQCIMIYEKNRYSISLDHLWKSNINIYFKKILLF 580

Qy         67 ------NFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120
                     |:  | |      :: ||:        |:  |        :
Db        581 LLLLLLFSSYNNLNNTNVLQLYD------CYKYNVHNN-------------------- 613

Qy        121 QYAAGDIFTFLNIYDKDNTKGIFISFYE----------IYCGKL--YDLLQKRKMVAA-- 166
                    :| :  |        |  :   |:|:            |:   ||    ||:  || :|
Db        614 -----YLFIYFNTLIVANRRYTFVSYSSSYALMRTINLIHLNKLYIYDIFIKRNLVGCHG 668

Qy        167 -------------LENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDES 212
                          :  |     :|:   :  :|  :  :     :::  |    ::  |:  |     |
Db        669 FTLFIHEKHETNDYIVNNNYNIVI--VLVLILVGRRNRTIQIEDRRIIYKLPVIILNLPS 726

Qy        213 SRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKE 272
               ::       ||    :  :  ||  |       ||               :|       ||  :  :|
Db        727 RKNKIYTNIYSCEKHFNTKL-YCFFI----------FKQRNNNEIVDGVHNEKS------ 769

Qy        273 CIRAMDSDKN-----HIPF---RDSELTKVLRDIFVGKSKSIMIANISPTISCC-----E 319
                        ||      |:  |     :  |   |  :   |::     |:          |   ||  |       |
Db        770 --------KNYYMNDHVKFVKNKKSFLIQTYSDLY--HFKNYNDNKIFNNISLCYNINLE 819

Qy        320 QTLNT----------------------------------------LRYSSRVKNFKNK 337
                :|                                            |  :   |  ||   |
Db        820 HRMNNVQKYIHADMSIHFKKNYKRDKGINQYIYNYKHNFYLRVFNFLLNFIKRTKNVINI 879

Qy        338 STCINEEDDTNTERISILDSKGSEMNASSIENVVI---------------------KSN 375
              ||         :: :  ||      |:      |  |:   |                              |
Db        880 VLCI--------FKLLILT---KEILVYYILNLFIYYYIWLKCYFVLYNIIYEYIERKDY 928

Qy        376 HLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKK 435
               |:       ||:|   |  |  |||  :           |  ||                         :|::  |:  |  |
Db        929 MLMGLYNNSK--RFKDNDKKD----LFNK-------------------NINEEKERKAK 962

Qy        436 GLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNI 495
                  :     :  ||||          ||||:||||||:|||||||:|||::                |:   |||
Db        963 FVSCATNNEYNDN------NNNNDNNNDNNNNNNDNNNDNNNNND-------NNNDNNND 1009

Qy        496 NNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKN 555
              |||  |  |||||||||||::||:     |     |  :|               ||||     ||
Db       1010 NNNNNYNNNNNNNNNNNNNNNNY-NNNNNGGDRND-----------ENNDKD----KKE 1053

Qy        556 LRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMED 615
                    ||  :  |    |:   |||               :     :     :    |    ::|:   :
Db       1054 DNDNDNINNMLENDD--GKKKRTRRRKTSDKSKTINKMTTGRRGRKSTSMLNAHID--AN 1109

Qy        616 NTPKDILYESRN---VSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNV 672
               |   |:  ||    |     ::|:|  ::||    |       ::      ::|  :  |  :  |  :  |  :|
Db       1110 NILKNNLYNINNGNYMNNINNSILLNNNMQENY----MNDNKSSNMLMNNIPNKMNMKMN 1165

Qy        673 NSINNSNMNSINNSNMNSNSIYKSNYN------------SNQSISDVQIRYV-NEMDTSN 719
                |||   ||  :  |||:|       ||                 |  ||:       ::  |  |  :|
Db       1166 GQINNEIMNDMINYNMNNNMGNHMNNNMGNHMNNNMGNHMNNHISNHMNNHISNHM--NN 1223

Qy        720 KNNDNIFF----------DAISCDNNMYP------NITNNNNNNNNNNNNNNNIDVENY 763
               :  |:||:               ::::   |:|  |            |:  |||||||:||||||   :   |
Db       1224 QLNNNIYVNNRVGIGYMSNSVNQHNHMSPHMNNTINVHNNNNNNNNSNNNNNNNNNNNN 1283

Qy        764 NNRDGT-----------NNSMKLYAYNSHNLFQPDNNKNTSNI----QNINTNKNNQDGN 808
                 ||  :  :                    ||  ||        |:  |  |        |  |:  ||           |||  |  |
```

```
Db       1284 NNNNSSSGSNCVNNTLRNNLMK----NNFNQFNRMVNNNSINIHTSSSNINNNMN----- 1334

Qy        809 VNYSMNFCHYNLNDKNYLIDLNNKEQKDK-NIHGCDNNI---------------IQNRN 851
              ||:::  |  :|  |  |: ||     | | |   | :                | | |
Db       1335 -NYNIHLTHQKINSNNNNINNNNLNMNGKLNYHQMSNKVHGTVPIPMNVHMGGTSISNIN 1393

Qy        852 DFEKKKKTNFYNNNNIVIVNNNMG----NNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMK 907
                      |  || | :   :| ||    |:  |  | |   |:   |  |  : |
Db       1394 HMNNINNMNNMNNINNMGSSNRMGTSPINHPIPHKKGGTHMSNQITRNAMNKMLSNPINV 1453

Qy        908 DNEMK-------------DNHIKSNNNNSSSSSSN-NNIYNNINDDDTFQ-------- 944
              ||:|                :| : || ||   :|:: | |:: |||  |
Db       1454 DNKMNVLNRKNSCTVNFNMNNPMVSNMNNYMNSNNMNRNSLPNNICSGGTIMPNVMSKGK 1513

Qy        945 -NDYCHNDNTFTIRRKNNTNINS------------------NIYQNDDIIYTINSLNDY 984
                  |   |:|:|   :|| |||:                  |    |  :  : :|::|
Db       1514 IKDTAINNNSF----RNNNNINNMNNMNSMNHMNSMNHMNSMNHMNNVNHMNNMNNVNSM 1569

Qy        985 MSNTL---------------------------LHFKEKYT---YP---------TLST 1003
              |:| :                           |   ||    ||          : |
Db       1570 MNNRMSEQNIIANNMIKTEMRNNNLMNMKLVNRNQHMNMKYVSEKYPMNTKMIQNNNMVT 1629

Qy       1004 NEDI------------------------YNKEMEG------------KHIRLDDQDKY 1025
              | |:                        ||  |:            |  |   | |
Db       1630 NRDMNIINNNNINGINNMGFINYMNKNNTLNKNMKNIKTENYTYDHHLLNHDPLSPMDTY 1689

Qy       1026 D-DNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNN 1084
              |  : |||:  || ||: |||:| |  ||:  |  ||::| |          |::    |
Db       1690 DIQHMNNNI--NNMNNI-NNVNNINNINNI--NSINNINNMN--------SIHSV----N 1732

Qy       1085 NEYLSYFQKNVDTIINNCLNSLDISS--MYDDTKEILNNILLSKYKAEKDNVIKKYINED 1142
               :::   |::  |:   | |||: :: |    |   |||       : : :    |: ||: |
Db       1733 TKFMLNKQEDDKTLHPNYLNNQNVQSNTYYHQTKEEEINTYNNTHSSTTINMKKKHKN-S 1791

Qy       1143 IKNMSLEEID-----KTAQSIYEKRKVLLTKL--------LLLFKKNV--DTQINNE--- 1184
              :||   ::::      :: :  :  |   |  ::         :| :||:  :  ||:|
Db       1792 LKNSKEDDLNNNQNLRSVKINWNKYDNLFKRIPLERNYEGFILEQKNIKDEDTINDEKTF 1851

Qy       1185 ----TSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQI--------WCESE 1232
                      |     | : | : | |      |  : :   ||:|: |  ::          ||:
Db       1852 MLENNEDKEDDAIYCDLSNIN------FKTFFKDYNDILNKLNLIKLNDKTNFKNNCYSQ 1905

Qy       1233 NLRLLYQ--FLVVE--------------------------YQNKSANSVLLNVSSNNG 1262
                ||:   ||::|                          ||||  |: | |:   |
Db       1906 ----LYEEIFLIIEKFLDISHLKYHNNFTHTNFYNNCLKYAIYQNKE-NTNLHNIKEENV 1960

Qy       1263 DIILLNKKL--------VQDNIKNSMDHN 1283
              |    || :          ::| |  |||
Db       1961 D----NKNIHGLDIHLTHENNETNKEDHN 1985
```

RESULT 7
Q8MMY3

```
ID   Q8MMY3     PRELIMINARY;      PRT;  1422 AA.
AC   Q8MMY3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG5792 protein. 6/101.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116965; AAM33216.1; -.
DR   InterPro; IPR002110; ANK.
DR   SMART; SM00248; ANK; 2.
DR   PROSITE; PS50088; ANK_REPEAT; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat; Repeat.
SQ   SEQUENCE   1422 AA;  157732 MW;  69A3D85CB1102363 CRC64;

  Query Match             12.4%;  Score 850.5;  DB 5;  Length 1422;
  Best Local Similarity   23.9%;  Pred. No. 8.9e-27;
  Matches  310;  Conservative  170;  Mismatches  358;  Indels  459;  Gaps   45;

Qy        206 NSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINR 265
              |: |: :: :::| | :     || | :: :         ::   |    || :|     |:
Db        261 NNNNNNNNNNSINNNNSNSINNNLNINR--------NKNDEDESFQNIET------IDG 306

Qy        266 SLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTL 325
              ::::  |     || :|     |    ::  |   |  :  |:  |   |
Db        307 NIVSADEV--NMDIEK------QSNSLNIIPDTLV---EEIIDPNTPPA---------- 344

Qy        326 RYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNK 385
                    |   |       |:|      :|||           :  ||  ||  |   |:  :|||||
Db        345 -------NHLNHQACVN---NTNTM---------NNMNNMSINN---NGNNNNNNNNNN 382

Qy        386 INRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLY 445
                 |    |:    ||  |                                  |   :
Db        383 SNNNNNNNSNNNNNNNSN----------------------------------NNNNNNN 407

Qy        446 NDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNN 505
              |:|   |    :|||||||:|:||:|||:|||||||::::  ||   |:  ||| |||  |:||||
Db        408 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNINNN 467

Qy        506 NNNNNNNNSH----------------NNHLPQP-------------------------- 522
              |||| |||  :                :| :  |
Db        468 NNNNKNNNGNGEYQQYIFDSEDHIIWSNEIVVPIEVSKTSKRSNSSINGSNGSSRKKKLK 527

Qy        523 --------NYAFTDTSDFS----------------SLDDMNCHLNNNDKSIFLHK----- 553
                      |  |:|:  |                ||    :    |    |   | :
Db        528 TSPTTATTNTTSTNTNTLSKNGGSPNIGPNSQPTNSLSSSSSSNNTQSSSEILQQLMNGD 587

Qy        554 ----KNLRDNIKLKNRS------------------SCDNIMN----------------- 573
                  ||  |  || | |                  :|  ||
Db        588 LDTIKNYVDG-KLSNDSKLVFDLEKKDEDGFSPLHYACAYGMNGFVEYLLRIRVSPNCPD 646

Qy        574 ----------KKKNNL------------HLARHSVGSKLTMFSYDP---------QKNK 601
                        | |||             :  ||    : | :||          |
Db        647 KEGWTPLHWSVKANNLDAVQLLLEFGADQSVPNHSRQYPVEM-AYDKGFTHIAATLQGVS 705

Qy        602 DNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLG---------------------- 638
               :  | |:| |  |:::   :     :    || | :|
Db        706 EYEFKKNNENNPENSSMISVGINGVSGGGGNGGVPMGGSSSSSSSSSSSSSSSSSSSSMM 765

Qy        639 ---------------------------LNKNTHHDISTKD---------ENHNDNKI-- 659
                                          | ||       :   |          | ||  |
Db        766 MVMMNNQGILEDSGGNGTAGGLSQEYRILLKNGLEPYTPPDGTNLKSGVIEKHNTLKFIQ 825

Qy        660 ------------------------------------------------NNGVINIINNSNV- 672
                                                              | |:  :::| ::
Db        826 AAEKLGLYSEHRSALMEGFKSLRTLRYRLFWIMLSNPLFSKICEWDNGGIDFLVHNKDIL 885

Qy        673 -NSI--------NNSNMNSINN-----SNMNSNS-------IYKSNYNSNQSISDVQIRY 711
               |::         :|| :: :::     | :||       :  :|| :: ::
Db        886 LNALGLYFGLRGSNSAIDWMSHLARRKCNKSSNSGGDVWTNLRASTFNSQTNLEKLKDIT 945

Qy        712 VNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNN 771
                : | | ||: :  | : | :||  |  |||||||||||||||   |   |    ||
Db        946 SEKNDKSRKNDKDDMFISHSFNNN---NNNNNNNNNNNNNNNNNNSSNNNNGNNGNNNN 1002

Qy        772 SMKLYAYNSHNL-----FQPDNNKNTSNIQNINTNKNN-----QDGNVNYSMNFCHYN-- 819
              :        |:·               :|| | :| |:|:|       :: |:|| :: |
Db       1003 NSNNNGNNNGNSNGNSGGNYNNNNNNNNSNNNNSNKSNNSNHDKNSNINYDNDYSGMNNL 1062

Qy        820 -------LNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNN 872
                     |: :  |  ||    : | :  :||   | |:          |  ||||  |:||
Db       1063 KSSNRRRANNNSNNIQNNNNNSNNNNNNSINNN---NNNNNNNNNNNNNNNNNNNIINN 1119
```

```
Qy     873 NMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSS----- 927
Db    1120 NIQNNSNSNSN---SNSNININNNNNNNNNNNNNNNNNNNNNNINNNNNNNNNSNYNNY 1176

Qy     928 SSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSN 987
Db    1177 SNNNHSNNNHSSSKNFKNSH-HNN----INNNNNNNINNNINNINN-----NSNNNNNNN 1226

Qy     988 TLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDN 1047
Db    1227 N-------------SNNHQNSNNNNNGNNHQNNNNNNNGNNSNNGIINNNNNNNNNNNN 1273

Qy    1048 NNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTI-------IN 1100
Db    1274 NNNNNNNNNNNNNNNNNNNNNNNNN----NNINNKNKN-YASDVSMNIDIIEEINSQTIK 1328

Qy    1101 NCLNSLDISSMYD-------DTKEILNN---ILLSKYKAEKDNVIKKYINEDIKNMSLEE 1150
Db    1329 NINNNTEINENSEIEEINSQPTQEQINHNEAMLISEEQSHQSD------NSDQDNAHLNN 1382

Qy    1151 IDKTAQ------SIYEKRKVLLTKLLLLFKKNVDTQI 1181
Db    1383 QQTTSQQHGSSNSISDTNQLILNESSPEFSNPEDIQI 1419
```

RESULT 8
Q8I548

```
ID   Q8I548      PRELIMINARY;      PRT;   1681 AA.
AC   Q8I548;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFL1815C.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014850; AAN36449.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1681 AA;  192792 MW;  3985DBE6957CB8A6 CRC64;

  Query Match             12.4%;  Score 847.5;  DB 5;  Length 1681;
  Best Local Similarity   24.8%;  Pred. No. 1.4e-26;
  Matches  329;  Conservative  223;  Mismatches  406;  Indels  369;  Gaps   63;

Qy     170 GKKEVVVKDLKILR-------VLTKEELILKMIDGV--LLRKIGVNSQNDESSRSHAILN 220
Db       3 GFDEMKMKQLQLLKEKMMKNAKAQKSEEVKKSVESVSNILNVSNVSNVSNVSN----VRS 58

Qy     221 IDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQT---DGANINRSLLALKECIRAM 277
Db      59 VDSKDIRFSVKSGKKIFL---------DTSFQNKKLMVHPKNGRNI-------------- 95

Qy     278 DSDKNHIPFRDSELTKVLRDIFVGKSKSIMI-----ANISPTISCCEQTLNTLRYSSRVK 332
Db      96 --------FQTSKNTS-----FVSKNKNSLLKLTTEEELFDTPLSFEEVMRRKNEKKKME 142

Qy     333 NFKNKSTCINEEDDTNTERISILDSKGSEMNAS----SIENVVI----KSNHLLSNNNNN 384
Db     143 LQKNNQTNISEE-FINNKQGSINQNSPNKNNSHLFKIQYDNLNISQDNQNGNTTNNNNNN 201

Qy     385 KINRGKIN-----------DKIERNNILKNKSFDKPREGFT-----------STFGKYS 421
Db     202 NNNNVNVNVNVNVNVDNNTHINNLQNNSLINGNLLQNNNSSYNVDTIKNTNSLNINSFKNFF 261

Qy     422 SLNDIDKIKKNKKKGLIN----YKSTLYNDNTINKKHNN----NNNN---NNDNNDNNND 470
Db     262 TQNISSNNEKNCNVGMLNLTFPFNNSCKDEKILTTKSVNGTVFNNINIVKNDDNNDDNND 321

Qy     471 NNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTS 530
Db     322 EDNNNNN-------------NNNDEDN---NNNNNNNNDEDNNNNNNL----------- 353

Qy     531 DFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNR---SSCDNIMNK-KKNNLHLARHSV 586
Db     354 -----------LN---KSTFFNKDSIFGTALLKNDYIFSTDEKYNNKDKENNI------- 392

Qy     587 GSKLTMFSYDPQKNKDNTFFKSNINKMEDN---TPKDILYESRNVSNMNGNVLLGLNKNT 643
Db     393 -------------NESHNIFKNDKNNMNGTNIFSTNNFMTSEKIISNKLNNTIINNNNQT 439

Qy     644 HHDISTKD---ENHNDNKINNGVINIINN-----------------------SNVNSINN 677
Db     440 QNNGMSQNSFTSNHNVDQVHNNTITSLNGTSAGIQKVNVDNICIPNKGPTNINNVNNVNN 499

Qy     678 SNMNSINNSNMNSNSIYKSNYN----------SNQSISDVQIRYVNEMDTSNKNNDNIFF 727
Db     500 VNNNIFNNSAVNNGGIFSSGKNLFGIPIARTTNNLFLQKIESNIINTNKMNNNNNNLMFN 559

Qy     728 DAISCDNNMY--PNITNN--NNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNL 783
Db     560 KTIDSSTNIFGKPCITNNIFNNTLGNNHINKNNNTLFNNINTNNNMNNII------SNNS 613

Qy     784 FQPDNNKNTSNIQNINTNKNNQ--DGNVNYSMNFCHYNL---NDKNYL-IDLNNKEQKDK 837
Db     614 FNPLITNNTLK------NEDNKVVVSNINHIENVIQFNKSSENDENNLNVSYKSVENKES 667

Qy     838 NIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMK 897
Db     668 TMTIGDNNNVFSFYNFKANKLTNLTNHNNNDNNNNNNDNNNN---NY----------NNN 714

Qy     898 NNEMKNNEMKDNEMKDNHIKSNN-----------NNSSSSSSSNNNIYNNINDDDTFQND 946
Db     715 NNNYNNNNNNDNIVVGDHLKSNSECVQGQNMNEQNNEMVYNKQNTNIIINTNNNDNVNKE 774

Qy     947 YCHNDNTFTIRRKNNT----------NIN-SNIYQ-----NDDIIYTINSLNDYMSNTLL 990
Db     775 II---TTMESPSENNTICKGEDHIKPQINVSNMDEVAKSGKEDTQHVINN-NINESKEPK 830

Qy     991 HFKEKYT------YPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNN 1044
Db     831 NIKKTFTSYFNNFIPSIFTSE---NKK-EPKDQQNGNNDNNDNNDNNENNKNNENN-ENN 885

Qy    1045 VDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHN-----IKN---FNNNEYLSYFQKN-- 1094
Db     886 ENNENNENN-ENNENNKNNENNKDKYDDVNKFENRTGDIITNDDSLTNN--LCIFKKNET 942

Qy    1095 VDTIINNCLNSLDISS-------------MYDDT------------KEILNNIL---LS 1125
Db     943 VENIPNNNDNNNDSSSNNKEEECLNNKNQLYMNTIKQDGGNSNVWNNSQNNNNFLFNNLS 1002

Qy    1126 KYKAEKDNVI--------------KKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLF 1173
Db    1003 GNHPNNDNILASKEVKPFLGSSEKTKNEPIELMTPNKINENSNTFFINKNVNNTNIL--- 1059

Qy    1174 KKNVDTQINNETSDLR------KDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQI 1227
Db    1060 SKGEEIE-NNENAEINLFNRIPKNGVF-NLSSENYENKEEFDKFNKIYTNIN-------- 1109

Qy    1228 WCESENLRLLYQFL---------VVEYQNKSANSVLLNVSSNNGDIILLNKKLVQDNIKN 1278
```

```
Db       1110 -CINKNINFINTYVPSNVVNINNFIEDQNSCANT--YDALSKTNDMFL--KKMKSDNNNN 1164

Qy       1279 SMDHNNI 1285

Db       1165 NNNNNNI 1171


RESULT 9
Q8I2P8
ID   Q8I2P8      PRELIMINARY;      PRT;  3178 AA.
AC   Q8I2P8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Protein kinase, putative.
GN   PFI1280C.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22255708; PubMed=12368867;
RA   Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA   Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA   Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA   Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA   Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA   Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA   Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA   Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA   Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA   Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA   Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch E.,
RA   Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA   Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA   Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA   Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT   "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL   Nature 419:527-531(2002).
DR   EMBL; AL929358; CAD51942.1; -.
KW   Kinase.
SQ   SEQUENCE   3178 AA;  366959 MW;  82FB35F0914C4DED CRC64;

  Query Match            12.3%;  Score 839;  DB 5;  Length 3178;
  Best Local Similarity   23.6%;  Pred. No. 5.4e-26;
  Matches  332;  Conservative  210;  Mismatches  420;  Indels  444;  Gaps   55;

Qy        196 DGVLLRKIGVNSQNDE-SSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNK 254

Db        207 DFLLNNKI-INSHNKEIIHTTNDVLNEHIMSSNINNSISKSNNNYYFGNEK---YDKDNN 262

Qy        255 QTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPT 314

Db        263 YTHHDDHFYNSSSYTNEGAIKKLKN--NPVLLRDS------------------------ 295

Qy        315 ISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMN---ASSIENVV 371

Db        296 ---CNENINTFDNINSNDNNNNNNNNDNNNNDNNKNNI----NEHTHMNSNISTNVSDIT 348

Qy        372 IKSN----HLLSNNNNNKINRGKINDKIERNNILKNKSFD----KPREGFTSTFGKYS-- 421

Db        349 ILENKKLLNILNINNNIENNFNSNNYSYDKNGLFESKDKDMNCLTPQEEICKVQDKENYK 408

Qy        422 -SLNDIDK-----IKKNKKKGL-------INYKSTLYNDN-----------TINKKHNN 456

Db        409 DFLNVINKRMINFTKDNNSNNININNIDNVNTKSSLFNDKIFMDNHTDNIMVTNNNNNNN 468

Qy        457 NNNNNNDNNDNNNDNNNNNNN-----------------------DSSSMVNNMIN---- 488

Db        469 NNNNNNNNNNNNNNNNNNNNNNYNSSCSSSSNKATNIYKYNTNQHSDSNKNTNNIVNINHE 528

Qy        489 ----------------------------------------------HMINNNINNNI 499

Db        529 YYKKLTPLKVKNENDEDEDIKLEGDYPNDQVIDNVPNDDYDDDNNQTHRHKSNNNNDNNN 588

Qy        500 NVNNNNNNNNNNNNSHNNHLPQPN-YAFTDTSDFSSLDDMNCHLNNNDKSIF--LHKKNL 556

Db        589 DNNNDNNNNNDNDNININDNDNDNCYGYIYKNKSSTLDESLKHQEVN-TSIFSYTNKAGI 647

Qy        557 RDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDN 616

Db        648 YNNNDVINFFNDDNIEMEYKN-----------ISSLSYNKNVCDASTNYNDNIN----- 690

Qy        617 TPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSIN 676

Db        691 -----------NNTNNNTSNNTNNNTNNNTSNNTSNNTSNNTNNYTNNNTSNNTSNNTN 738

Qy        677 NSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSN---------------- 719

Db        739 NNTYNNIRVSNNNHCDDHHNSYDTENIINYEDLLYNNDLKEKNSDHFIDVEEGIREYNYK 798

Qy        720 ----------------KNNDNIFFDAISCDNNMY----PNITNNNNNNN-----NNNNN 753

Db        799 RLLSYQKIQTNESEDIRKNNDNYYMN--NLNNDMFSSNDENIKTNEPNNSISSFYNKTNS 856

Qy        754 N---NNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVN 810

Db        857 NFLINNDICDDNKKDQPSVPEKNRHISYNSYEKEEEQNLKHSNSLENNYINVENARINNF 916

Qy        811 YSMNFCHYNLNDKNYLIDL-----------NNKEQKDKNIHGCDNNIIQNRNDFEKKKK 858

Db        917 FS---CNYTLDDKTKINNRNTTYTYNGNNNNNNNNIDDDKNDHGDDKNNNIDLNNYKKRNN 973

Qy        859 TNFYNNN-----NIVIVNNNMGNNNSPRMKYGLCGSHTSIDNM---------------K 897

Db        974 SNNSSNNIYHYDNKVYMFDSVKKNISTIMNYEPYSYHSKMNNVIKDEHAFYGSFLNKEKM 1033

Qy        898 NNEMKNN----------EMKDNEMKDNHIKSNNNNSSSSSSSNNNIY-----------N 935

Db       1034 NNKIKNHKNNNMNKSYIDIHDTDTNDNFISDDFNNNMGEDNYVNNVYDDEDDVNNNVDGN 1093

Qy        936 NIN------------------DDDTFQNDYCH---------NDNTFTIRRKNNTNINS- 966

Db       1094 NIDGDNIDGDNIDGDNIDGDNIDGDNIDGDNIDGNNIDGNNIDDNNIDGNNMNSNNMNSN 1153

Qy        967 NIYQNDDIIYTINSLNDY-------------MSNTLLHFKEKYTYPTLSTNEDI---YN 1009

Db       1154 NMSESNSVNISVHDLKNVKYNSSINLEDQKKCMPNSSYHDIKANTYEKINHNMDVRVHSN 1213

Qy       1010 KEMEGKHIRLD---DQDKYDDNDN---------------------------NNVDN-N 1036

Db       1214 NNHKTNHDNMKNVFDQHCYVSNGNMEYIKDNMNTVNSINSMNSMNSMNSLNDVNNVNNIN 1273

Qy       1037 NKNNVDNNVDNNNVDNNVDN-NDKNNVDN-NNVDNDDDDVDFHNIKNFNNNEYLSYFQKN 1094

Db       1274 NVNNM-NNVNNMNNVNNMNNVNNMNNVNNMNNMNNMNNMNINNINNFNN---------- 1322

Qy       1095 VDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDKT 1154

Db       1323 ----VNHS-NSMKSIGEYNMNNYNLHNIRMS--KGDDNNTNKVDVGNHVNN--------- 1366

Qy       1155 AQSIYEKRKVLLTKLLLLFKKNVDTQINN---ETSDLRKDLVMCHICNNNPDDQFHFYAY 1211

Db       1367 SQIIGMKKNDVIYSDMSMNYENMNNNINNYNINNSNNHNDLL-----PNNINDQYN---- 1417

Qy       1212 SRLEKDII----NLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDI--- 1264

Db       1418 ----NDSLKFKNNFIMSNNL-CGSTNMN---------------NINILNNVNNLKDINTM 1457

Qy       1265 ILLNKKLVQDNIKN---SMD-HNNIH 1286

Db       1458 DLLNSSRKYNSMKNVVSTMDILNNVH 1483


RESULT 10
```

```
Q8MN35
ID   Q8MN35      PRELIMINARY;      PRT;   828 AA.
AC   Q8MN35;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117082; AAM34352.1; -.
DR   InterPro; IPR003034; SAP.
DR   Pfam; PF02037; SAP; 2.
KW   Hypothetical protein.
SQ   SEQUENCE   828 AA;  93950 MW;  673800D237DF142E CRC64;

  Query Match             12.2%;  Score 838;  DB 5;  Length 828;
  Best Local Similarity   28.6%;  Pred. No. 1.7e-26;
  Matches  281;  Conservative  140;  Mismatches  322;  Indels  240;  Gaps   45;

Qy        206 NSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADT-VSQNKQTQTD----- 259
Db         14 NNNNNNNNKYNNYNFD------NTSNGEVKFCSKCKSSVAYETPVDMNKLNEESIRKIL 67

Qy        260 ---GANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDI-FVGKSKSIMIANISPTI 315
Db         68 GYYDKNCHNGLLC-NDCFNGF--IKNRIN-ENKKLTEITENAHFLGR------------ 110

Qy        316 SCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSN 375
Db        111 ----EPLQPFKIN---KNFNNNN--FTDESPNN-----------------YSNCVIAST 143

Qy        376 HL--LSNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNK 433
Db        144 PLNSINNNNNNNNNNNNNNNNNNNNNNFVYSHNYN----------------NNIPHIKIDN 187

Qy        434 KKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNN---------NDNNNNNNNDSSSMVN 484
Db        188 ----INNNSNHNNQNGNNNNNNNNNNNNNNNNNNNNNNNVDDSKTNNNNNNNNNNNNNNN 243

Qy        485 NMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNN 544
Db        244 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSLDEGK---------SKLKNMCLFYLN 294

Qy        545 NDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFS---------- 594
Db        295 SLKTVQLHELIDRTN---SSKKGC-----KAKLTERLLRH-------MFAFTRTIEIPPL 339

Qy        595 YDP-QKNKD-----NTFFKSNINKMEDNTPKDILYESRNVSNMN--GNVLLGLNKNTHHD 646
Db        340 YTPSMKDSDIQMHLKRYGAQTHGNRSEMIPRLLVYHLLFFDRFNPDGTPVVGNNNN---- 395

Qy        647 ISTKDENHNDNKINNGVINIINNSNVNSI----NNSNMNSINN------SNMNSNSIYKS 696
Db        396 ------NNNGNNGNNG-----NNNNNSSTQRTPRSSPRPSTNNTKPSPRSQLNSATLNSS 444

Qy        697 NYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNN 756
Db        445 NPNSSTSTTTTTTTTSNTNT-NINNNN-------------STTNNNTNNNNNNNNNNS 489

Qy        757 NIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFC 816
Db        490 N---NNNTNNNGYNNS-----YNNY----VDNSLNVS---------NSEVSNASPSTPYT 528

Qy        817 HYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGN 876
Db        529 GYDLNNSGNGI-----------VNGGYNNSYNNNNNNNNSNNSNTYNNNNI---NNNYNN 574

Qy        877 -NNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYN 935
Db        575 YNNYYNNNYNYTGG-----NSQNSFNLNNYNGYNTINSNNFVNNNNNNINNNNYNNNSRE 629

Qy        936 N-----INDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLL 990
Db        630 NSSPVLVPTNDKIQ--IC-VDSGLCEGLAYNNGYNHSIYNNKDYAATITPLN-YQPTVEL 685

Qy        991 HFKEKYTYPTLST---NEDIY--------NKEMEGKHIRLDDQDKYDDN-DNNNVDNNNK 1038
Db        686 QQNQDYYHQFLSTFDGGNNLYYIPTPTHQTQQPNNYYYYVNGNNNNNNNIGNNNINNNND 745

Qy       1039 NNV-----DNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQK 1093
Db        746 NNINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNINYNNN 805

Qy       1094 NVDTIINNCLNSLDISSMYDDTK 1116
Db        806 NNNIAIDNNINYNNINFINNSKK 828

RESULT 11
Q8IDV7
ID   Q8IDV7      PRELIMINARY;      PRT;  3256 AA.
AC   Q8IDV7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF13_0210.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844509; CAD52512.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   3256 AA;  384176 MW;  57724DD1ABD5220E CRC64;

  Query Match             12.2%;  Score 838;  DB 5;  Length 3256;
  Best Local Similarity   23.9%;  Pred. No. 6.1e-26;
  Matches  348;  Conservative  217;  Mismatches  462;  Indels  430;  Gaps   64;

Qy         16 ELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDTVDNFTVYENTI 75
Db       1880 ENEEKDEENEEENDEEN-----DEENYKKDNEEDYNQHKNCNDQFMKIEGDN--IFDNST 1932

Qy         76 KPLIIDLYENGCVCSCFAY--------GQTGSGKTYTMLGSQPY--GQSDTPGIFQYAAG 125
Db       1933 SNSLHNI--------CYSYKNRYNDSHAYLMSGENEIMAHNKIYRPQEKDIP----ITQD 1980

Qy        126 DIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRVL 185
Db       1981 NSFVFQN---KNN------------GNQYD---------------------------- 1995

Qy        186 TKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSER 245
Db       1996 ------AQMSDFYNFHLNSMNSINDTNS-----IN-NMNDINNMNGINNMNGINNMNNIN 2043

Qy        246 GADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDI----FVG 301
Db       2044 NLNNIN-NMNNMNNLNNINN--------LNNMNSINNIHNMNNMNSINYTNNINNTNSVN 2094

Qy        302 KSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSE 361
Db       2095 NMNNINNMNNMVTIN----NMNSINNMNNIQNVNNLYNFYNE----NQNNISVPNIHGEF 2146
```

```
Qy        362 MNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYS 421
Db       2147 MSEKTNEE-KISTNDLDTYEN---LNKDMIND-LSINMSLNNYNNEMMNNEICSSY--FD 2199

Qy        422 SLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSS 481
Db       2200 MCNNMDE----------SY-NIMNNNNSTNSSNNNNNNKKNENENENNENNNNINNNNND 2248

Qy        482 MVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQ-------------PNYAFTD 528
Db       2249 NNNN------NNDNNNNNNDNNNNNNDNNNNNNNNNNLFECDKCNVDYDIMNFTNNTYMD 2302

Qy        529 TSDFSSLDD-MNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVG 587
Db       2303 IQKILQNDDILKQEFANIEECFKIHTSSI-NYMKVKGKGECCH-----KNQQIMYEH--- 2353

Qy        588 SKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDI 647
Db       2354 -------YDAGKHSNNNFIRNN----------DDMYKNLSMDE------FFLNPN----- 2385

Qy        648 STKDENHNDNKINNGVINIINNSNVNSINN---SNMNSINNSNMNSN------------ 691
Db       2386 -----NIEANIINNDIINENINSDIEKIMDIMTSNNNNEREDNKNHSRQEDDKGRQQTKG 2440

Qy        692 ---------------SIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNM 736
Db       2441 KEKCKEKDKYKDKEFDIYK-EINNDKMISEHMGIYGEQINLKKKE-----FTNFSSSSSF 2494

Qy        737 YPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNN----------SMKLYAYNSHNLFQP 786
Db       2495 --NMSNNNNNNNNSNNNNNNNNNNNYNNNNNNNNHTFDISTKNKDKKRKIDSLGNTYEH 2552

Qy        787 DNNKNTSNIQNINT-------NKNNQD--------------GNVNYSMNFC-----HYNL 820
Db       2553 IISSNESKIESINEQSTSRKFHKINEDKRKKYQKLDIRTLNKNVQKNKEICTNCYIHYDN 2612

Qy        821 NDKNYLID-LNNKEQKDKNIHGCDNN------IIQNRNDFEKKKKTN--FYNNNNIVIVN 871
Db       2613 FKSSYILTFINRKQKKQRKIFPVEPNEKDEAYIVQIIKYIEKLKEQDKIFGINKNEEKQE 2672

Qy        872 NNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSN------------ 919
Db       2673 NKNNCLNEGEQQQGVLNINTNINNCNDNIMKENNMKDNIMKENIMKGNEVAHHDSMNDHK 2732

Qy        920 ---------NNNSSSSSSSNNNIYNN---INDDDTFQNDYCHNDNTF---TIRRKNNTNI 964
Db       2733 VINKDESNENNMNNNNLNNNDNIFNNDNIFNNDNIFNNDTIYNNNIYNNDTI--YNNDNI 2790

Qy        965 -NSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQD 1023
Db       2791 YNNNIYNNDNI---YNNSNIYNNNNIYN------------NSNIYN-------------- 2821

Qy       1024 KYDDNDNNNVDNNNKNNVDNNVDNNNVD---------NNVDNNDKNNVDNNNVDNDDDDVD 1075
Db       2822 ------NNNIYNNNDNN--NNCSNNNVNMMNQFPYGNNTEYILNNKLTFNNNDLNQDKID 2873

Qy       1076 F--------HNIKNFNNNEYL--------------SYFQKNVDTI--INNCLNSLDIS-- 1109
Db       2874 LINDMSRTNENIKNNNMNNFVQNFNEQFFVQQTNFNYINDNINNISPINNMNNINGLKKH 2933

Qy       1110 ---------SMYDDTKEILNNILLSKYKAEKDNVIKK----YINEDIKN-----MSLEEI 1151
Db       2934 MNRTYSRSISMYEDI-NIPHNVDYINYSIENDNVYDENYNIVLNQCASNNANVTMFFESN 2992

Qy       1152 DKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFYAY 1211
Db       2993 DNKDNSTKPSGDVYMEDM------NMPNNMNNMNS--------MNSMNSVNSINSINSVNS 3039

Qy       1212 SRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKL 1271
Db       3040 INNGNNINNLNIMNNI-----------------------NTLSCNINNNSFANNLYDNNI 3076

Qy       1272 VQDNIKNSM---DHNNI 1285
Db       3077 ISHNMKNIMYPYINNNI 3093
```

RESULT 12

```
Q8I350
ID   Q8I350      PRELIMINARY;      PRT;  3069 AA.
AC   Q8I350;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFI0495W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22255708; PubMed=12368867;
RA   Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA   Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA   Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA   Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA   Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA   Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA   Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA   Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA   Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA   Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA   Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch E.,
RA   Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA   Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA   Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA   Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT   "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL   Nature 419:527-531(2002).
DR   EMBL; AL929356; CAD51785.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   3069 AA;  368341 MW;  731D5B1C386643B3 CRC64;

  Query Match             12.1%;  Score 830.5;  DB 5;  Length 3069;
  Best Local Similarity   21.6%;  Pred. No. 1.2e-25;
  Matches  374;  Conservative  230;  Mismatches  483;  Indels  643;  Gaps   72;

Qy         21 KKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDTVDNFTVYENTIKP--- 77
Db        614 EKDEKYLNNYNNTCSY-EQNKYMID-TNYLYNHE---DK------NNYSTNKNDLNERNK 662

Qy         78 --LIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPY-GQSDTPGIFQYAAGDIF---TFL 131
Db        663 RFSSKNLVNNSHISEC------------------PYTDEQISKFLFIHMNNNILKDEKKL 704

Qy        132 NIYD----------------------KDN--------------------TKGIFISFYEI 149
Db        705 TIFDDEKSDDDHHHHDHNNNNNNNNIRDNNYLNNHVSLNSNNFVSHGLKNEGFKNNLNNI 764

Qy        150 YCGKLYD----LLQKRKMVAALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGV 205
Db        765 ECQKINDQSDKLFYDNRQIMCPYNKKKDIFNNDVKKLLNLKNDNL--------------- 809

Qy        206 NSQNDESSRSHAILNIDLKDIN----------KNTSLGKIAF----IDLAGSERGADTVS 251
Db        810 NYSSYSSNSIHRKYSDDINSLHIKKSICKHSCNNIKKGHMKYNCYNHDMLYNEHEKRSML 869

Qy        252 Q----------NKQTQTDGA---NINRSLLALKECIRAMDSDKN---------------- 282
Db        870 EYYSRVAKRNCNRLIQKDNVRSYNINKSIHRNERINNIYNNNNNCDNNNCDNNNRDYHCD 929

Qy        283 --HIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTL-------NTLRYSSRVKN 333
Db        930 TRHYKDSSNEYNK-----YNGNDNICNDHNLSHTLSKYDNFVLSDNNNDNSNNYSLKEEN 984
```

```
Qy        334 F---------------------KNKSTCINEEDDTNTERISILDSKGSEMNASSIEN-- 369
Db        985 FNFNLIINKEKKRPPKLSNNKNNENYKNCENYENYENHDNYENYENYDNYENHKNYENHK 1044
Qy        370 ---------------------------------------------------VVIKSNHLL 378
Db       1045 NYENHKNYENHKNCENAMYNDLFQGEINQNGKYVHDKECFKNDEYIFKNSKNLLKNNELL 1104
Qy        379 SNNNNNKINRGKINDKIERNN---ILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKK 435
Db       1105 FKNNESLF---KNNENIFKNNENLIKNNENIFKNNENIFKNNEYIFKNRNLEFLKNDKKK 1161
Qy        436 GLIN-YKSTLYNDNTI------NKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMIN 488
Db       1162 KFQNIFYANCPEENYLNMKLIDNQKKKNNNDNNNNNDKNNNDNNNNDHNND--------- 1212
Qy        489 HMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKS 548
Db       1213 ---NNNDKNNDN-NNNDNNNNDNNNDNNN-----------------DD------NNIEN 1244
Qy        549 IFLHKKNLRDNIKLKNRSSCDNIMN----KKKNNLHLARHSVGS----KLTMFSYDPQKN 600
Db       1245 FFNHFKD-------------DNMVNIFSDRNNSNYHATKISYDMNKLYKGTRLSY---KE 1288
Qy        601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHN--DNK 658
Db       1289 KHQTTNEQKVNM--------ICYEKNNNTENN------NKNNNKNNNTENNNKNNFDNN 1333
Qy        659 INNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTS 718
Db       1334 NNNNNINNYDNHNINNCDNNNKNNYDNHNINNCD--KNNFDNNNN------NNINNYDNH 1385
Qy        719 NKNNDNIFFDAISCDNNMYPNITNNN-NNNNNNNNNNNNIDVENYNNRDGTNNSMKLYA 777
Db       1386 NINN---------CDNNNKNNYDNHNINNCDKNNNNNINNYDNHNINNCDNNN----INN 1432
Qy        778 YNSHNLFQPDNNKNTSNIQN---------INTNKNNQDGNVNY-SMNFCHYNLNDKNYLI 827
Db       1433 YDNHNINNCDNNKNKLNKNSLIEEYCNYKIHIHKDERRNAGEYLSLNRQKLSVSKKIEII 1492
Qy        828 DLNNK-EQKDKNIHGCDNNIIQNRNDF-EKKKKTNFYNNNNIVIVN--NNMGNNNSPRMK 883
Db       1493 KKNNEINYKRKIIDDNSRNIISNNYYFPEFNKSSLFYKNNNLYNISKKNNLHGMN----- 1547
Qy        884 YGLCGSHTSIDNMKNNEM------KNNE-------------------------MKDN--- 909
Db       1548 -----KFNNVRNLKNNEINHLFFQKNNQLQIRRSIQIDKNKETNFEHVTDFDKLKENKFI 1602
Qy        910 ------EMKDNHIKSNNNNSSSSSSSNNNIYN-NINDDDTFQNDYCHNDNTF-------- 954
Db       1603 IYNNKLQLINNHNNNNNNNNNSNNNNNNSNNINSNSNNNVSSDLHHKNSVISYNNMDFN 1662
Qy        955 ---TIRRKNNTNINSNIYQNDD------------------------------------- 973
Db       1663 QNNTIMKKNKNSVLFNCYHNEDEEEKHKYIKKNVIKNELDKNNENNTEKIIDYIIDVDKI 1722
Qy        974 -----IIYT--------------------------------------------INSLN-- 982
Db       1723 YDNKYVIYDNNLLQLQRKKKKKVENYDFVLDIDKECTPRNKLLLEKLNKNIAYINALELS 1782
Qy        983 ---------DY--MSNTLLHFKEKYTYPTLSTNEDIYNKEMEGK----------HIRLDD 1021
Db       1783 SNKYHNRSYDYYVQKNFLHHLQEKENITRYNKNDYIYKYDIQEKRKKKKNRGKSNIYIHD 1842
Qy       1022 QDK-----------------------------YDD--NDN-------NNVDNNNKNNV--- 1041
Db       1843 QEQFMFNTNDFIKNRTNSIHSHYLRDENKLIYDHQFNDNNTPYMLYNNLSRRNRNSMNYH 1902
Qy       1042 ---DNNVDNNNVD-------------------------NNVDNNDKNNVDNNNVDNDDD 1072
Db       1903 HDEENNEDEEYLDIEEDDEDEHHHQYVNENYYRHKENNKNNNNNNEKNNYNNNNNNNN- 1961
Qy       1073 DVDFHNIKNFNNNEYLSYFQK----NVDTIINNCLNSLDISSMYDDTKEILNNILLSKYK 1128
Db       1962 --------NNNNNYYFCHNNKTEMEKITEIQNMLSKKIRLGNVNDENYNTNQPIFNKKFS 2013
Qy       1129 AEK---DNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNET 1185
Db       2014 FEDFDLDNFITNSMDLDNKD---ENIIKNILEL-ERRKYAINCIMEKLRKD-----QNES 2064
Qy       1186 SDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEY 1245
Db       2065 EYLSDDSDQHHVLNLKMMNTHIYDLYS--DKSTYNVY---------NTKMCINYLLEEK 2112
Qy       1246 QNKSANSVLLN--VSSNNGD-----IILLNKKLVQDNIKNSMDHNNIHKK 1288
Db       2113 MHSEENKEFIKNFFVENNADHINFMEIFKKKKKKKYNHKNNNSHNKENKE 2162
```

RESULT 13

```
Q8ICX0
ID   Q8ICX0      PRELIMINARY;      PRT;  2874 AA.
AC   Q8ICX0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein, expressed.
GN   MAL6P1.48.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
RA   Berriman M., Pàin A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844505; CAD50319.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2874 AA;  334409 MW;  65541406B59D8DCB CRC64;

  Query Match            12.1%;  Score 830;  DB 5;  Length 2874;
  Best Local Similarity  23.6%;  Pred. No. 1.1e-25;
  Matches  323;  Conservative  213;  Mismatches  412;  Indels  422;  Gaps  56;

Qy        206 NSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQN-------KQTQT 258
Db         33 NSRDNPNDRDN-LNNNDNPNINDNPNNN--------DNPNNNDNPNNNDNPNNNDNPNNN 83
Qy        259 DGANINRSLL---ALKECIRAMDSD-----KNHIPFRDSELTKVLRDIFVGKSKSIMIAN 310
Db         84 DNPNINEQYLNNFLTKEGDNYIDEDLLVYTKNEFLYAYKEL--ILDKKGLGEDHELPLCS 141
Qy        311 ISPTISCCEQTLNTLRYSSRVKNFKN------KSTCINEEDDTNTERISILDSKGSEMNA 364
Db        142 -----------NKFKFPELEKNIRNFDEDNDQQYNVSFERDFDNRRLNDEHDFNNEKNR 189
Qy        365 SSIENVVIKSNHLLSNNNNNKINRGKINDKIE---RNNILKNKSFDKPREGFTSTFGKYS 421
Db        190 YKNRNEFFYNNN--NNNNNNNYNNNYNNERNDFYCRNNL--NDSFD------------- 231
Qy        422 SLNDIDKIKKNKKK--GLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDS 479
Db        232 SLNNNDNNNMNRRSNFSIRNNKFNKFHDNNMSNLNNNNNNNNNNNHNNNNNNNNFDHNNR 291
Qy        480 SSMVNNMINHMINNNINNNINVNNNNNNNNN------------------NNNSHNNHLP 520
Db        292 NSYKKS--NYATDNSNNNNNNLNNNNNMSGTKKLLWRDSDKTGSMSTWRLGKNSVSNH-- 347
Qy        521 QPNYAFTDTSDFSSLDDMNCHLN--NNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNN 578
Db        348 -----------NSSMNHMNNNMNKLNNTNMMMSHNNN-----KMSDNKRGHKTFNKNYNN 391
Qy        579 LHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLG 638
Db        392 --DAKSNISSDM----------KNSTSHEINVDNM--NNYMHHLYNNNDMTKMN------ 431
```

```
Qy     639 LNKNTHHDISTKDEN---------HNDNKINNGVINIINNSNVNSINNSNMNSINNSNM 688
Db     432 -NKNNY----TKDRNIIHMKREDYYTNESHYNDG---LNNTNNINHMNNN--NNINNNNN 481

Qy     689 NSNSIYKSNYNSNQSISDVQIRYVNEM----------------DTSNKNNDNI---FF 727
Db     482 NNNNKHMNNFGNN------QLPYIQHMNNNESGAGGAGGPNDIVADHSKRNVENFNSPSF 535

Qy     728 DAISCDNN---------------------MY--------PNITNN------------- 743
Db     536 VSVNKNGNIQNNSSNYYNKGVNNVTKKGSEMYNTSYYGLSPNINNNMDDWKMKKYKKFSN 595

Qy     744 -----------------NNNNNNNNNNNNNI-DVENYNNRDGTNNSMKLYAYNSH--- 781
Db     596 NNMNSMNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNSNAHYNNHHHL 655

Qy     782 ---------------------------NLFQPDNNK----NTSNIQNINTNKNNQDGNVN 810
Db     656 NNAGVHESYRLIEEKFKKDSIQNEYPNNLYMNKNNEMNKMGPNSINLIRDNNNNNNNNFN 715

Qy     811 YSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNI------IQNRNDFEKKKKTNFYNN 864
Db     716 YNNNF-NYNQHQYGQKGSINVNVTNNINTHNNANNTFNYMNNVTNNNDVVEKNKGYYKDS 774

Qy     865 NNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKD--NEMKDNHIKSNNNN 922
Db     775 SNKNMNNNNIGNNNT----------TVFENMENNLAKKSEKSNTPSHHSFDNAKNQTTN 823

Qy     923 SSSSSSSNNNI-----YNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDII-Y 976
Db     824 FTGSHYYNKNLRTNEKRNSKNEKSVIEDDWS------LIRKK--SSITDNKKASDDLLWH 875

Qy     977 TINSLNDYMSNTLLHFKEKYTYPTLSTN-----------EDIYNKEMEGKHIRLDDQDKY 1025
Db     876 KMSEQFDRNNKSILELSEKKKDQVTNKNDNKPEAKNKNKESSKDKKNEGCNVNNKKDEKN 935

Qy    1026 DDNDNNNVDNNNKN-----------------------NVDNNVDNNNVDNNVDNNDKNN 1061
Db     936 NNNNNDDPDNNNNNNSNYNTGTTPLTNTAMTGVKGKNTNSNNNNNNNNNNNNNNNNNNN 995

Qy    1062 VDNNNVDN-----------DDDDVDFHN------------IKNFNNNEYLSYF------ 1091
Db     996 NNNNNKSNRMYGMATSSNIKEHPVDRNNYNKDKIMNNNMSVLKNKKNNE--SVFTNGSSN 1053

Qy    1092 ------------------QKNVDTII-----NNCLNSLD-------ISSMYDDTKEILN 1120
Db    1054 INSVNNMFKENEMNVEKKKKKNVELNIHGEKNDNQLNSKDGNNSPSKLNSNQNNNNNNN 1113

Qy    1121 NILLSKYKA--EKDNVIKKYINE-DIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNV 1177
Db    1114 NNKQKKKKKKNEKDNKLTKNVEDNDVNKKLYNALDNMASELKQKD-------ILLNGNDE 1166

Qy    1178 DTQINNETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLL 1237
Db    1167 KTNNNIHNNESHKSYV--HENYVFGDNQINNYYWNKQQNNNITYEL------DHNNNDNL 1218

Qy    1238 YQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKLVQDNIKNSMDHNNIHK 1287
Db    1219 MQKNIMLMDNIPQDSLLQKIKNNNMSVENLNNALLNTHVNKNMEMMNISK 1268
```

RESULT 14
Q8I5I9

```
ID   Q8I5I9      PRELIMINARY;      PRT;  2432 AA.
AC   Q8I5I9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFL1085W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014847; AAN36303.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2432 AA;  284064 MW;  EF0989EB11BD3A33 CRC64;

  Query Match             12.1%;  Score 827.5;  DB 5;  Length 2432;
  Best Local Similarity   21.4%;  Pred. No. 1.2e-25;
  Matches  361;  Conservative  223;  Mismatches  407;  Indels  699;  Gaps   66;

Qy     204 GVNSQNDESSRSHAILNIDL-----------------KDINKNTSLGKIAFIDLAGSERG 246
Db     564 GYMEENNFILNSHSSNSLDMGTHICYNYQNKINHNITKQCNDNN--------------- 607

Qy     247 ADTVSQNKQTQTDGA----NINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGK 302
Db     608 -NTVFHNNYSHKDNVKVLNNIYNNIITDKE----RESNDN-----DSTKQSMLNDDDADK 657

Qy     303 SKSIMI-----ANISPTISCCEQTLN---------------------------TLRYS 328
Db     658 LDNIKLNNYIYSNMNDYFTICDPNINIDKTTLHENYLRFQPNVLEKDVNGGNKHMTMRY- 716

Qy     329 SRVKNFKNKSTC------INEEDDTNTERISILDSKGSEMNASSIENV------VIKSNH 376
Db     717 --IDNFKHNSISNHNNNNNNSNNNNNSNNSNNNNNSNNSNNSNRLYNISSEALNSLKKND 774

Qy     377 LL-----------SNNNNNKINRGKINDKIERNNILKNKSFDK---------------- 408
Db     775 SLYMRLQNIISFSSGHNNTKIKVININNFNNNANKFKNLNMDNMNAPNGVTNLNVGINNE 834

Qy     409 -PREGFTST-FGKYSSLNDIDKIK--KNKKKGLINYKST----LYNDNTINKKHNNNNNN 460
Db     835 VVKEEKTNTCLNINEQIHSHTKLKCIKNGKPRNINQKENNVVYIESDNSHNDNNNNNNNN 894

Qy     461 NNDNNDNNNDNNNNNNNDS--SSMV---------------------------------- 483
Db     895 NNNNNNNNNNNNNNNNNNNSFENNMLLLCKSKNVKNQNRMNNLFRKEENKDTKLKSLNYNI 954

Qy     484 --------NNMINHM--------------INNNINNNINVNNNNNNNNNN------NNS 514
Db     955 LINPCYNKNNVINELFMNTKEKSLYGNTNKFHNTLNNNMNMNNYNNQNINNIYPLTHNIS 1014

Qy     515 HNNHLPQPNYAFTDTSDFSS-------LDDMNCH--LNNNDKSIFLHKKN---------L 556
Db    1015 IVNHILNHSYYNENTEQNNKNLFHNTILVPKNCNNIMNNSDKYNYLNAYNQNYTLYNNTL 1074

Qy     557 RDNIKLK-----------------------NRSSCD-------------------NIMNKKK 576
Db    1075 NNNNSLVQPSDVNNMNCSYESINGNMPINDQSCDTYNDEYNDTYNDTYNDTYNDTYNDTH 1134

Qy     577 NNLHLARHS---------VGSKLTMFSY-DPQKNKDNTFFKSNINKMEDNTP-------- 618
Db    1135 NDTHNDTHNDTYNDEYNDTYNEINNETYNDTLNDTPNDTLYDTLNGITNDTPYITPTNAC 1194

Qy     619 ----KDILYESRNVSNM-----NGNVLL--GLNKN--THHDISTKDENH----------- 654
Db    1195 NLSYEELMYKMQNNNNIYCKKHDSNYIISNGINNHICAHKEVNYNVLNHVTNINTYVNSK 1254
```

```
Qy        655 ---------NDNKI---NNGVINIINNSNVNS--INNSNMNSI----NNSNMNSNSIYKS 696
Db       1255 NIFSNVAVTNDTDLHACNNEQRNNISNHNINKNVVNLNNVHSINKELNNYPINSTQDYNT 1314

Qy        697 NYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDN-NMYPNITNNNNNNNNNNNNN 755
Db       1315 NYDKNGCLGTLTCDIKNKNNKNNKNNKNNKNNKNNKNNKNKNNNDNNNNNNNNNNNSNNN 1374

Qy        756 NNIDVENYNNRDGTN----NSMKLYAYN-------------------SHNLFQPDNNKN 791
Db       1375 KDIIILNKNNINNANIFNINSLKEKLLNDDISKNKEKKKKKCVIRKIKKRNKVNPKRNNN 1434

Qy        792 TSN----------------------------------------IQNINTNKNNQDG--- 807
Db       1435 INNRNKNNGSYKCDTKRVNKINNGFMKNFKNKYTSSCINYNMCAISDLSTNIKNKENFVI 1494

Qy        808 -------------------------------------NVNYSMNFCHYNLN---------- 821
Db       1495 GNYPYTNGVSYITNGENNTTNNGKKNDDHRNKKKEYNSNENNNYCNNNSNRNNNNNNNS 1554

Qy        822 ----DKNYLIDLNNKEQKDK----NIHG------------------------------ 841
Db       1555 HFGGNKNNRGDDNDEDDQEKNNKRNIYGKNGDDEEEDYDGEDDNNEENDHNNKKNIKRKN 1614

Qy        842 -----------CDNNIIQNRNDFEKKKKTNF-------------YNNNNIVIVNNNMGNN 877
Db       1615 KKSNKAYNNNDDENKCINNKDTFHLPKKHNYEYTMNNNYNMNNEYNKDDEEIKNNYEDNQ 1674

Qy        878 NSPRMKYGLCGSHTSIDNMK-----------NNEMKNNEMKDNEMKDNHIKSNNNNSSS 925
Db       1675 DNGDIKHKSCTNGNDVDGVKSKYHKNDHSGDNNNDINNNSDSNNNNNNNNNNNNNNNNN 1734

Qy        926 SSSSNNNIYNNIND----------DDTFQ-------------NDYCHNDNTFTIRRKNNT 962
Db       1735 NNNNNNNNNNNNNNNSCTNNCATVDDLMNLNSSNLGKFSYLTNVLSHNIN--NINFINNN 1792

Qy        963 NINSNIYQNDDIIYTINSLN--DYMSNTLLHFKEKYTYPTLSTNE-DIYNKEM------- 1012
Db       1793 NHKMNFLNNENVNQFFNNMNYKDYVNN--MNYLEPNNYINHNINELNYFNKNMILNNLNY 1850

Qy       1013 ----------------------EGKHIRLD-------------------DQDKYDDNDNN 1031
Db       1851 INYVNGINNDRYMNGINNSQMSYNNHFRNDIISNVLKGVRVINKESHYKNKNNKNGNSNN 1910

Qy       1032 NVDNNNKNN-----VDNNVDNNNVDNNVDNNDK-NNVDNNNVD----------------- 1068
Db       1911 NNNNNNNNNNEDEDDDNNQNNNNDNDGDDNNDNDNNEENNNIDDRIPFKHHNMNNVSMKN 1970

Qy       1069 NDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYK 1128
Db       1971 NDGKNEDNNNNKNNNNN--------NNNNNNNNNNNSNDNNNNRNGNNNIINPFFLYNKM 2022

Qy       1129 AEKDNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDL 1188
Db       2023 FCMNNNANNFNNGNI-------------------------------------INNFTNN- 2044

Qy       1189 RKDLVMCHICNNNP--------DDQFH-FYAYSRLEKDIINLIMLRQIWCESENLRLLYQ 1239
Db       2045 ----GMAH--NHNPNMNINSFRNNNMHLFHKYGIQYKDLIPMNMFPTAY-----LNLGRE 2093

Qy       1240 FLVVEYQNKSAN--------------SVLLNVSSNNGDII--------LLNKKLVQDNIK 1277
Db       2094 RRSSKYENSSYNKKKDLMKRRYELQKAILLNVDENLKEVIDEIIRNSTILPQKGIRG--R 2151

Qy       1278 NSMDHNN-IH 1286
Db       2152 NTLDCNHPIH 2161
```

RESULT 15
Q8I5A3
ID Q8I5A3 PRELIMINARY; PRT; 1824 AA.
AC Q8I5A3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Asparagine-rich protein, putative.
GN PFL1530W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014849; AAN36392.1; -.
SQ SEQUENCE 1824 AA; 212658 MW; 6ECED0EA6960084E CRC64;

```
  Query Match             12.1%;  Score 826.5;  DB 5;  Length 1824;
  Best Local Similarity   23.2%;  Pred. No. 1e-25;
  Matches  376;  Conservative  244;  Mismatches  450;  Indels  549;  Gaps   76;

Qy          8 VVRKRPLSELEKKKKDSDIITVKNNCTL-YI----------------DEPRYKVDM---- 46
Db        191 ILKDNPLNEQNDNKSN-------NVCTLNYIKKKTSLDNDHSRIIKNNEPKEHINQEALV 243

Qy         47 ---TKYIERHEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTY 103
Db        244 RSKNPYAESKLEKIEQLFEEIKKQEDLYVDEIGRLLLLSFDNLC---------------- 287

Qy        104 TMLGSQ----PYGQSDTPGIFQYAAGDI------FTFLNIY--DKDNTKG---------- 141
Db        288 TLLVDQIKDVKLGSKD---IEMFLRGEIEKIKDNNNMLNRYLIKMQNEKNDMLFNMNDNM 344

Qy        142 ---IFISFYEIYCGKLYDLLQKRKMVAALENG--KKEVV--VKDLK----ILRVLTKEEL 190
Db        345 LNDIKISFYDIF----------EKML----NGEYKKEFEEHIKEDKNKRIIYIILCKKQI 390

Qy        191 ILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTV 250
Db        391 IYILEE--LYKNVYKNKIKDE------------KYINKLYNMNKFMYI------------ 424

Qy        251 SQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIAN 310
Db        425 ----LNKANNNSFNNNIV-----------EKKNHI---------IVNHVENLKRETEQVES 461

Qy        311 ISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERIS-ILDSKGSEMN--ASSI 367
Db        462 VLENV--IENVLDDEEQSFEKKEKINKIMDRNFLEDKYVEKIDRIHDSDIFENNHVLSND 519

Qy        368 ENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDID 427
Db        520 TNVSLESN---ASNKTNQINKLNLNKQIS-NCFSKDNNNNNETQFSLSTCLTESSFMD-K 574

Qy        428 KIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDN----------------------- 464
Db        575 KMKENSSK-----ETNITNENQNGEDYSNDILNNNDNMNNVHCDNSTFMKIHYNDQFNNH 629

Qy        465 -----NDNNNDNNNN--NNDSSSM---VNNMINHMINNNINNNINVNNNNNNNNNNNS 514
Db        630 YPITINNNNNDSNNNNMYNNMSNNIYHNISNNNNNISNNNMNNNMNNNMNNNTTNINNNN 689

Qy        515 HNNHLPQPNYAFTDTSDFSSLDDMCHLN-----------NNDKSIFLHKKNLRDNIKLK 563
```

```
Db        690 QFNHTHFNNNLINQYNDQYNI-PLNNHLNNQLTTQGNCNMNNELS-----KQLSNQFNNQ 743

Qy        564 NRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDP----------------QKNK------ 601
              :|  :||:   |   |       :    |: |          .     ::||
Db        744 VKSPLNNIVGSANN---LVNEYFSDMMVYISWVPKSARCQYPRELPKSAAERNKLSEYIV 800

Qy        602 -----------------DNTFF-----------------KSNINKMEDNTP---KDILYE 624
                               || :|                  :|  |   |||  ·|: ::
Db        801 AKEQMLYILRVMGYDEIDNIYFHPPKGSHIKIKFKSMICMNNFLKAYKNTPDKWKEDMFR 860

Qy        625 SRNV--------------------SNMNGNVLLGLNKN-------------THHDIST-- 649
              |:                      |        :  ||||              |: |
Db        861 FFNIPLRSSISVRDLRIERAVPRSSAAEFKKMKKLNKNYQDLFLFKDNFNVCEEDMDTIE 920

Qy        650 ---KDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISD 706
                 | :  | |    | ::   |:  :| :  |  : ||::|    : :|:  || :|
Db        921 NMEKLDFHYDTYNKNVILQNKNDITMNKLLNEDKNSLSNMKNANINIHDENYPNNN---- 976

Qy        707 VQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNR 766
                   :| :: :||||: |   |             :   ::|||||||:|||||  |
Db        977 -----INNINNNNKNNNIILLD---------EHHHHHNNNNNNNSNNNNNAI-------- 1014

Qy        767 DGTNNSMKLYAYNSHNLFQPDNN-------KNTSNIQNINTNKNNQDGNVNYSMNFCHYN 819
                |||: :     :: : |   |         |    :  |  ||| |  | |    |  |  :
Db       1015 --VNNSLSIPVAGINSYYAPGVNVHGFLGSNNAVMMNNTITNKTNNDLN-----NLCSSS 1067

Qy        820 LNDKNYLIDLNNK--------------------EQKDKNIHGC----------------- 842
               |||: |:||:                      ||||
Db       1068 TGAKNYMKDINNRLFVGTKIDNNTLNINTDHFNTDVMKNIHNTLTTTTTTMINKNENDI 1127

Qy        843 --DNNIIQNRNDFEKKKKTNFYNNNNIVI-----VNNNMGNNNSPRMKYGLCGSHTSIDN 895
                ||:   |   |   :   ::   |||||  |          ||||  |||        : : :|
Db       1128 MNYNNLNHNNTMFASRTSSSRNNNNNININSNINVNNNNNNN----------VNNNNN 1177

Qy        896 MKNNEMKNNE--MKD-------------NEMKDNHIKSNNNNSSSSSSNNNIYNNIN-- 938
              : ||    ||     :|:                 |  :   |:   :|||:::::  ::|||  || |
Db       1178 INNNNGSNNMYLLKNKSLNRNGSILSNLNFLNTNNYNNNNNDNNNNINNNNNINNNNNML 1237

Qy        939 --DDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDII-------YTINSLN-DYMSNT 988
                         |      :|  |:  ||   ||||:||   |::|:        :|: |  |:::|
Db       1238 SSASVNLNNANLNNINSNTI--NNNTNMNSTNIMNNNILSNNILSNNNVNNTNKDFLANN 1295

Qy        989 LLHFKEKYTYPTLSTNED----------IYNKEMEGKHIRLDDQDKYDDNDNNNVDNN-- 1036
               :  |:    |  :|::|:  |                ||   |     ::: :   :   |  ||    ||
Db       1296 -FNLKKNTHYASLTSNGNNMNLLNEQFISYNSNTE----NINNMNSNNMNSNNMNSNNMN 1350

Qy       1037 ----NKNNVD-NNVDNNNVDNNVDNNDKNNVDNNNVDN------------DDDDVDFHN 1078
                   | ||:: ||:::|:::|  |:: ||:::||:::|                 ::  ::  |:
Db       1351 SNIMNSNNMNSNNMNSNNMNSNNMNSNNNHINNNHINNHHINNHHINNHHSNNHHINNHH 1410

Qy       1079 IKNFNNNEYL-SYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKK 1137
               | |    |   : ::   |||    :|   |::||:      ||   ||:||: ||
Db       1411 INNHPGNHNMNAHADKNV-AYVNPLNNNIDINS----KKE--NNLLLANYK-------NS 1456

Qy       1138 YINEDIKNMSLEEIDKTAQSIYEKR----------KVLLTKLLLLFKKNVDTQINNETSD 1187
              |:|          |      | : : :             :| || : ||  |        :|||   |
Db       1457 YLNN--TNSGSNTFSNCGQRVNQLKCLNTYNNKLNEELLNEELLNEK-----YVNNE--D 1507

Qy       1188 LRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESE-------------N 1233
              |: ||                           |::||:  :   |  |  |               |
Db       1508 LKDDL----------------------KNLINIDFINDIDMEYEEKHSYNNETFLRTSN 1544

Qy       1234 LRLLYQFLVVE-YQNKSANSVLLNVS--SNNGDIILLNK------KLVQDNIKNSMDHN 1283
                ||  :      :  :  : |::|  ||     ||||:|      ||   :|:  :|    :  |
Db       1545 SSLLQRHTYPKNFPSPNENTLLTVVSSTSNNGNINKMREYQTNYDKLPYNNVMNCDEQN 1603
```

RESULT 16
Q8IEE4
ID   Q8IEE4   PRELIMINARY;   PRT;   2672 AA.
AC   Q8IEE4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PHAT201.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844509; CAD52316.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2672 AA; 310742 MW; 8929A8A84AC08464 CRC64;

```
  Query Match             12.1%;  Score 826.5;  DB 5;  Length 2672;
  Best Local Similarity   22.5%;  Pred. No. 1.5e-25;
  Matches  368;  Conservative  215;  Mismatches  410;  Indels  641;  Gaps   72;

Qy        158 LQKRKMVAALENGKKEVVVKDLKILRVL----------------TKEELILKMIDGVLLR 201
              :::|| |: ::  || :|  |    |                           |||          |
Db          1 MKERKKVSKVKESKKSLVENDNFCLNNYRNNFYMENDNKKNMKEEKEE-----------R 49

Qy        202 KIGVNSQNDESSRSHAILNID----LKDINK--NTSLGKIAFIDLAGSER-----GADTV 250
                |   |:|::|     ::::|   : ||||  || |             : :  |           | :
Db         50 KEIFTSRNNDS-----LVDVDYNKMMNDINKYENTILN-------SNNSRCNNFPNYDQI 97

Qy        251 SQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIAN 310
                |      ::|  :|  |        : : ::: | :                        : ::| :|
Db         98 IYNNNVLSNGP-VNNS-------VNSNNTNNNMM------------------NNNMMNSN 131

Qy        311 ISPTISCCEQTLNTLRYSSRVKNFKNKSTCINE-------------EDDTNTERISILDS 357
                :          :| |  :: :   |    | :|:                  |:   :  :   :
Db        132 M---------MNNNRMNNNIMN----SNMMNDNMMNDNMMNDNMMNDNMMNDNMMNDNM 177

Qy        358 KGSEMNASSIENVVIKSNHLLSNN--NNNKINRGKINDKIERNNILKNKSFDK------- 408
                 |    :: |  : |::::|   ||| |  :|:  |:  |   |||:   :| :
Db        178 MNDNMMNDNMMNDNMMSNNIMNNNIMNNNIMNNNIMNNNIMNNNIINKSNFSEVTNYSTS 237

Qy        409 PREGFTSTFGKY-----SSLND-------IDKIKKNKKKGLIN--------------YKS 442
               :| |   |     |      ::||         |  |   |     :  |                        |
Db        238 NKEPFIQTQNMYLKDANNNLNSCYVNTSMCDSICSNMSNNVSNNMVDKITVDHSCMSNNS 297

Qy        443 TLYNDNT-INKKHNNNNNNNND---------NNDNNNDNNNNNNN-------------- 477
                :|:|:  |   :||||: |||           ||| ||| ||: ||
Db        298 NDHNNNSNCNYYYNNNNDYNNDYNNDYNNDYNNDYNNDYNNDYNNNYYNNTKGQEGQKQG 357

Qy        478 ---------------------------DSSSMVNNMINH--------MINNNINNNINVNN 503
                                          || : |||  |:        :|||: ||| | ||
Db        358 KIGNGKLVNMDNMFNINEHNNIYRERIDSLNTVNNTSNYKYNNYVNIINNDENNNNNNNN 417

Qy        504 NN-----------------------------------------------------NNNNNNN 512
              ||                                                       ||||||
Db        418 NNANYKNMENPPFKTVTNMKYAENLLKDSNDIHVVKNDVSLFNHTCDIVSNPYNNNNNNN 477

Qy        513 NSHNNHLPQPNYAFTDTDSDFSSLD------------------------------------ 536
                :|::    |        :||| | | :
Db        478 CGSSNNIIHSNPLVENTSSFCSENHTSIYPPKNKKNNYNNNNNNNDNKQLCNFNNTSYN 537

Qy        537 -----DMNCHL-NNNDKSIFLHKKNLRDNIKLKNRSSCDN------IMNKKKN----NLH 580
                      :|| |  |||: ::  :| |      |  :   | |:|:|                |      |      |::
Db        538 FPEQTHINCDLSNNNNYAVLTNKYNSNLNNESYNVSTCNNNNDFNHANNNNNNYNIQNIN 597

Qy        581 LARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVS--------NMN 632
                :   :|  ||  :    :     |  :||   |    ||      |        ||  :   ||                 ||:
Db        598 IKDINVNIKLENYFNEVLNNNNNT--KEYINNTYMNKHIDIDTQHMNVHLNNSIKSINMD 655

Qy        633 GNV------LLGLNKNTHHDIST-KDENHNDNKINNGVI---------NIINNSNV--NS 674
               |:             ::       || |    :    | |::|  :: ||                   || ::|:|  |
Db        656 NNLEHINFPMMETQPNTPHIFNNILNNNHSNNFESDSVISNIKEITRNNITDDSSVIRNL 715

Qy        675 INNSNM----NSINNSNMNSNSIYKSNYNSNQSISDV--------------------- 707
```

```
                 :||: |       |: ||:|   :| | | :    | :: ::
Db        716 LNNNQMDHKLVNNNNNNNYINNDIKKMDGGGNVNLQNMPKTMSIVTNAVTNSVTSSLGNV 775

Qy        708 ------QIRYVNEMD-------TSNKNNDNIFFDAISCDNNMYPNITNNNNNNN----- 749
                    :|:| |           |||::::: |        |   ||  ||||||
Db        776 LENIKRPSQYMNNMSYLGNVNKYINKNDEHLSF-------NKEDNIIGNNNNNNNAENTK 828

Qy        750 ------NNNNNNNNIDVENYNNRD--GTNNSMKLYAYNSHNLFQPD--NNKNTSN----I 795
                    |:|||:|| |  | || |  |  ||        ||:|:| :  | ||  ||    :
Db        829 NHNINSNDNNNDNNNDNNNDNNNDDSGNNN------YNNHHLNKHDKINNNFNSNWIRKL 882

Qy        796 QNINTNKNNQDGN----VNYSMNFC--------HYNLNDKNYLIDLNNKEQKDKNIHGCD 843
              ::   : :|:|  |       :|  : ||         |:|  :  : |:|:|     :|    |
Db        883 KSFGASSDNEDENEELPINANHNFYINNFLLNQHFNFNKNDNIRDINSKMSGGNSISE-D 941

Qy        844 NNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKN 903
              |:: |  | |:          ||||| |||| |||          :  :: :|:|| ||
Db        942 NDM--NHNNINS-------NNNNINSHNNNMNNNND--NMNNMNNNNNNINNMNNNNNNI 990

Qy        904 NEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDD-----------DTFQN------ 945
              | | :|     |:: :||||  :: :::|||| || |::                || |
Db        991 NNMNNNNNNINNMNNNNNNNMNNMNNNNNNINNNNNNNNHNNNSSNGKNMDTAMNSLNMLI 1050

Qy        946 ------------------DYCHNDNT----FTIRR------------------------- 958
                                :  | |      | :::
Db       1051 LPKIKGVRFDKSGRRWVASWSQNGNQKRQYFPVKKFGQAQAKYLAIYARIQAVNCLQRKP 1110

Qy        959 ----------------------------KNNTNINSNIY------QNDDIIYTINSLNDY 984
                                          |  | |:| |        ::||    |  : |
Db       1111 HKPRMSKRLSEKKKLLLNMEGKEKNELGNNECNNNNNNYISKNGDEDDD-----NIEDSY 1165

Qy        985 MSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQD-----------------KYD- 1026
                : :   :| |      :   |:|          :|  :|| |                   ||
Db       1166 NDDEICEDEENYV--DIDHNDD--------EHYDIDDNDDNGDNGDNSHNDDNDDEHYDI 1215

Qy       1027 ---DNDNNNVDNNNKNNVDNNVD-NNNVDNNVDNNDKNNVDNNNVDNDDDVDFHNI--- 1079
                  |||||| |||| || |||  |  |||  :|| :||| || |       :|       | :: |
Db       1216 TNNDNDNNNNDNNNDNNNDNNNDNNNNNNNNNNNNDNNNNDIKIMDQQSGDKEYVGILNN 1275

Qy       1080 ------KNFNNNEYLSYFQ-----------KNVD------------TIINNCLNSLDIS 1109
                    ||     |    :   |             |||:                :| ||    : |
Db       1276 ESKNVGKNIGQNIGQNIGQNVGKNVGKNIGKNVEKNIGPDIKDEYMKSITNNDTQTKDQD 1335

Qy       1110 SMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNM---SLEEIDKTAQSIYEKRKVLL 1166
                  :  || |  |   | :        |||         : ::|:      |  :  |  || |::
Db       1336 IANETTKIIYPNNITDNNHIVNDNVCLIMASNNLENIPEDKKHEQNVTIHSIDEQKD--Q 1393

Qy       1167 TKLLLLFKKN----------VDTQINNETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEK 1216
               :     |   :||              |         :  ||  :::        |  :  |:         ||
Db       1394 SNKFLRERKNKRMHNNYDNFYDNSFKKKLSDYNEEI------NKSSDE----------EK 1437

Qy       1217 DIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKLVQ--D 1274
              ::     |:   ||  ||               :  :|       :| |   :|   :||: |  :   :   |
Db       1438 EVSQTNSLK-YKCEEEN---------DQNYDKQERKILQN-DTNECEIICINLESTKQID 1486

Qy       1275 NIKNSMDHNNIHKK 1288
              |:|:     ::||:
Db       1487 NLKDL----SLHKE 1496
```

RESULT 17
Q8T806

```
ID   Q8T806      PRELIMINARY;      PRT;    901 AA.
AC   Q8T806;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 102.9 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115682; AAL92689.1; -.
DR   InterPro; IPR000629; DEAD_box.
DR   PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   901 AA;  102858 MW;  C3B695494117057B CRC64;

  Query Match             12.0%;  Score 824;  DB 5;  Length 901;
  Best Local Similarity   26.5%;  Pred. No. 6.8e-26;
  Matches  278;  Conservative  139;  Mismatches  289;  Indels  344;  Gaps   40;

Qy        417 FGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNN 476
               | |:|        ||:|| |  : :   |   :|:| ||   :|||||||:||:||: :| ||
Db          2 FSKFS-------IKENKPKKVKSNSSKGFNNNEINSTNNNNNNNNNNNNNNDSDNFINN 54

Qy        477 NDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLP--------------- 520
              |||       ||        ||| |||      ||||:|| ||          :|  :
Db         55 NDS----NN------NNNFNNN---NNNNSNNINNKTEDSNDISLLINQFIGSSPAPSTC 101

Qy        521 -------------QPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRS 566
                           |:        ||      :|    ||: |||  ||         |   :|        | :
Db        102 SGSDKFSIKPKKIVTPSKRKNSTSGSDIIDFNNCNNNNNSKSNSDSNNNKYNNSN-SNNN 160

Qy        567 SCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESR 626
              ||:|  |                |::  |   ||    |    |    :  |:  .                ||
Db        161 SCNNTNN------------GNERKMKGYDYGNNNSNNNYSSS---------------SR 192

Qy        627 NVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNS 686
              :                |:::: | :| |:|:|  ||         ||:| |: |||| ||   :
Db        193 DY------------NSNNNSSPRDYNNNNN--NN------NNNNNNNNNSNNNSPRDY 231

Qy        687 NMNSNSIYKSNYNSNQSIDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNN 746
              | ||         :| |:| |    |                       || ||:          :||    | |
Db        232 N-NSRDYNNNNTNNNSSPRD-----------SNSNNNRNSSSRDFSNNNSSPRDYNNSRN 279

Qy        747 NNN---NNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKN 803
              :||   |||||||      |: :  :|| :  | |:| |      | | |           : |
Db        280 SNNSFLNNNNNNNN------NSNNSNDNSPRDYNYSSSNSSPRDYNNNNDTF----SKKR 329

Qy        804 NQDGN---------------------------VNYSMNFCHYNLNDKNYLIDLNNKEQKD 836
              |  ||                            ||          |       |:||   |  |
Db        330 NNVGNSILEIKSNPTSPRIKPIMVWPRSSPRIVN--------KLPPVQSLVDLTEIEMID 381

Qy        837 KNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNM 896
                 | :  :||   | |:           |  ||||     |||    :::| :              ||
Db        382 DNNNNNNNNNNNNNNNNNNNNNNNNSNNNNNKSGSNNNFSDSDSDK-----------DNY 429

Qy        897 KNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTI 956
                 ||   ||          ||:  |||||:|:|:|:|:|   :|  |:::      |:   :|:|
Db        430 NNNNNTNNN-SFNSDSDNYNNSNNNNSNSNSNSNSNSNSNSNNNNNNNNSNNNNNNNNN 488

Qy        957 RRKNNTNINSNIYQNDDII----YTINSLNDYMSNT---------------------- 988
                 ||:| ::| :  |  |:       :|| ::| ::
Db        489 NNSNNSNKSNNNFSNIKILEKRRKSINEISDNLTTRNIEEIDSKMAHEFEHPFLEKLIFA 548

Qy        989 -------------LLHFKEKYTY-----PTLSTNEDI--------------------- 1007
                           |  :| :: |          ::     ||
Db        549 GQRPPFVECDDYFLKRYKNQFQYYLDISMKSIINTRDIGEGDDDDEINSVPFKCTLCSYE 608

Qy       1008 YNKEMEGKHIR--LDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNN 1065
              |::|    ::     :|: ||   :   ||| :||| || :|| :||| :|| :||: || :|:
Db        609 YDEENNNENTNGFIDEADKIYNRSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNS 668

Qy       1066 NVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNC-LNSLDISSMYDDTKEILNNILL 1124
              .   | :|:::: :|::  | | ||       |||  :               :    :::|:|     || |
Db        669 NSNNNNNN-NFYDKKYFKNN----YFQHLLSKHLNQIKAKGIQDTNVYNDEA----NIPL 719

Qy       1125 SKYKAE----KDNVIKKYINEDIKNM----SLEEID-----KTAQSIYEKRKVLLTKLLL 1171
```

```
                  | :      : ::|     : || :       |:||     |  :    :|| |  :|
Db        720 GSKKCDHLECSEIILKNEFTKHIKKLHHHHKDEDIDICPRCKILKKSIQKRIVKFSK--- 776

Qy       1172 LFKK-------------------------NVDTQIN---------NETSDLRKDLVMCH 1196
                ||                         |:|  ||        | |::|  |
Db        777 --KKQEKDDDNNSKPVSNKVATISHNPSPNNIDKDINNNNNNKNNKNNTNNLEND----- 829

Qy       1197 ICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLN 1256
                 || :   : |  :  | :|:                        |  ::  |  |:  |
Db        830 ---NNDNSSNNNYLENEEEPEIV-------------------FQDIKPNNNNNNN---N 863

Qy       1257 VSSNNGDIILLNKKLVQDN--IKNSMDHNN 1284
                ::|| |    |     ||   :|: |:||
Db        864 NNNNNDD----NNNDNNDNNDNENNNDNNN 889


RESULT 18
Q8IIQ7
ID   Q8IIQ7      PRELIMINARY;      PRT;  1483 AA.
AC   Q8IIQ7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Asparagine-rich antigen.
GN   PF11_0111.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014837; AAN35697.1; -.
SQ   SEQUENCE   1483 AA;  173923 MW;  F93DB19A2A34D5F1 CRC64;

  Query Match             12.0%;  Score 819.5;  DB 5;  Length 1483;
  Best Local Similarity   25.5%;  Pred. No. 1.6e-25;
  Matches  319;  Conservative  160;  Mismatches  333;  Indels  439;  Gaps   57;

Qy        189 ELILKMIDGVLLRK---------IGVNSQN-----------------DESSRSHAILNID 222
              || ||  | | : |         | |||:|                 ::   :: :  :|
Db         11 ELDLKNDDVVDVEKDVITNHAGNINVNSENGSNVEIKEGDEKDHIQENDKDETNVVNEVD 70

Qy        223 LKD------INKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRA 276
              |||       |||        :   | | ||         | | |  :        :
Db         71 LKDDEVTYENNKNE--------ENKSSNRWADMCEDCDAPLVDTYNDNNDMKG-----KV 117

Qy        277 MDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKN 336
              :|::      | :|::   ::  :      |:       :::  :| |  :  |
Db        118 VDNN-----FNNSDMNNNMKQNYYMAPKN----DMNNYATCYEIYIRNLHIDI------T 162

Qy        337 KSTCINEEDDTNTERISILDSKG------------SEMNAS-------------SIE--- 368
              |    |   |      :||:||: ||                :|| |             | |
Db        163 KEEIYNLFDGYRIKRINILNKKGKTAAAYVEFDNIEDMNRSLELNGKMIYSGNNSFEHGT 222

Qy        369 -NVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDID 427
               :|:|  : : : : | | |    | |||  | :|    |:|               |:::
Db        223 ISVIINKDKMKNFHVNQKQN--KFKKMINRNN---NNNFSNIRQG---------NNNMN 267

Qy        428 KIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNN-- 485
               |  |             || |  |   :|||||||: ||:: ::||||||  : | ||
Db        268 NIMNN-----------YNGNNNNNSNNNNNNNNSSNNNHMSNNNNNNNNMYNHMNNNNP 315

Qy        486 ---------------MINHMINNNINNNINVNN------NNNNNNNNNSHNNHLPQPNY 524
                             : : ||||:| | | ||        |||      :| | || | |
Db        316 FFNDNNKNMKMDMMKNVGNTINNNMNQNFNYNNIKGSYLNNNQQVGGSNVHMNH--QQNN 373

Qy        525 AFTDTSDFSSLDDMN-CHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMN-----KKKNN 578
                 |  : :; :  |   :||||      ::|:::  ||    |  ::  :|| |        :::
Db        374 MLNDKMNNNAFNFKNQNYLNNN----HMNKQSIADN-NNNNNNNNNINNTQSGTEQRKK 428

Qy        579 LHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYES---------RNVS 629
              | | : ||           :  :|      | | ::      |   |   |: || ||        | :
Db        429 LILKKRSVPLDQKSYVVNP-----NIFGEA---KPVDIDPEKILNESTTKEKDMNDTNSN 480

Qy        630 NMNGNVLLGLNKNTHHDISTKDENHND-----------------NKINNGVINIINNSNV 672
              | | |       | |:  |    |:||  |                           : :||   |   ||:|:
Db        481 NNNNN----NNNNSDDDAEKKNENIEDMEKKKEQQEGTDKYKKLSHLNN---NTTNNNNI 533

Qy        673 NSINNSNMNSINNSNMNSNS---IYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDA 729
               |   |   |    |:|:|:||    || | |:|:: :      | |      : |||:|
Db        534 NEFNKDN----KNTNMSSNKNVFNYKKNNNTNRNFN-----YNNNFANNTKNNNN----- 579

Qy        730 ISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNN 789
                        || ||||||||||||      |  :  | ||       ||:||
Db        580 ---------NINNNNNNNNNNNNNNNTKKIFMMNKNN---ANNNMK-------------- 612

Qy        790 KNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQN 849
              || :|:::  | :  |     |:|:|          | |    | : |:||
Db        613 KNMNNLESNNMSGNKNLANLNHS-----GNTNSNNMVNDMNN------------------ 649

Qy        850 RNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDN 909
                    || |    || |    | || |||                   |:|||     |:  :      |
Db        650 -------KKMNNVNNKNF---NKNMDNNNITR----------DFSNLKNNTPDDNKKNIN 689

Qy        910 -EMKDNHIKSNNN-----NSSSSSSSNNNIYNN-INDD-----------DTFQNDYCHND 951
                :  |||  |||      |     :|:|||: || :|::            : |:| : :  |
Db        690 VSVIHNHIYENNNKLYGSNQEDETSTNNNMNNNYVNNEMGTSGNMKGKFNNFRNTHGNMD 749

Qy        952 NTFTIRRKNNTNINS---NIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYP-----TLST 1003
                        |||          || : ||      :|  |   |  |    :    ::    |  |       ||:
Db        750 KKNVDHNNNNTPYQKKELNIIKKDD---KVNIKNIIMKN--IKREDNNTAPYNSGVNLSS 804

Qy       1004 NEDIYN---------KEMEGKHI-------------RLDDQDKYD------------DND 1029
              : :  |           | | |::                 :||:: ||             :
Db        805 SSNFRNDSSSNMRGQKMNEKKNVENDLLNNDRDGLKNMDDEEGYDNINRNMNSKGNTERY 864

Qy       1030 NNNVDNNNKNNVDNNVDNNNVDNNVD--------------------------------- 1055
              |||::|||  ||  |:  :  |:::|   |
Db        865 NNNINNNNNNNNNSSYNYNMENLYDSSLKYVKNANNLNKTKKITAVKTVKVIATEKKYD 924

Qy       1056 --NNDKNNVDNNNVDN------------------DDDDVDFHN--IKNFNNNEYLSYFQK 1093
                 | |:  | |:|| :|                   ||  :|  |   :|: :|   |  |
Db        925 DENGDEENKDDNNSNNIVNGETGNEKHYKDDKYHDDKKLDNTNESVKSNDNYTYESDVSD 984

Qy       1094 NVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSLE---- 1149
                 :||  ||   |: ||:|         : : :|  :| |        :||      :: :|  ::|
Db        985 TNNTIDNNNNNNKDITS-----TNLKSTLLQTKTKRR----VKKKTTDETENNNVENNEA 1035

Qy       1150 --------------EID-------KTAQSIYEKRKVLLTKLLLLFKKNVDT 1179
                            :||         |  ::|:|  |    |  ||:  |:|  |
Db       1036 NGDSNANAASGKSQKIDILIKPNLKKGENIWEAR----THLLVEQKENAVT 1082


RESULT 19
Q8ILS4
ID   Q8ILS4      PRELIMINARY;      PRT;  4466 AA.
AC   Q8ILS4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF14_0170.
OS   Plasmodium falciparum (isolate 3D7).
```

```
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014818; AAN36782.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   4466 AA;  519821 MW;  ECFA4865DCBFFA76 CRC64;

  Query Match             12.0%;  Score 819.5;  DB 5;  Length 4466;
  Best Local Similarity   21.9%;  Pred. No. 4.5e-25;
  Matches  399;  Conservative  263;  Mismatches  505;  Indels  659;  Gaps   82;

Qy         18 EKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVD------KVFDDTV---DNF 68
Db       1205 KSSKKNKNNINNKEGSSTFLDEKK-KSQLTENEENNMKSVDDNNNNKNNDNIVKGGNNV 1263

Qy         69 TVYENTIK-PLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAAGDI 127
Db       1264 TLLKNKKKGSVTYDLKNN---VEIGRYGKIQENNETKTDDDQPIGEN---SLFDNNFED- 1316

Qy        128 FTFLNIYDKDNTKGIFIS-FYEIYCG-----KLYDLLQKRKMVAALENGKKEVVVKD--- 178
Db       1317 YTYYTDYTDDPIMSDDISDDYE--DGEDEEEESDDRNDDNKYITGINDRYKEIKIEDDVQ 1374

Qy        179 ----------LK-ILRVLTKEELILKM--IDGVLLRKIGVNSQNDESSRSHAILNIDLKD 225
Db       1375 IRRDVFFFNLLKSVIKKKTSEDMLIYMNFLNKIKMEKIFDNIIQLSCEIFKTLLN--FIE 1432

Qy        226 INKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLL----------ALKECIR 275
Db       1433 ITKDSSTYRLLLKNL-GAWIGIITIGRNKPLMSKYMNIKQLILYAYDNGYLIVTFPALCK 1491

Qy        276 AMDSDKNHIPFRDS-----ELTKVLRDIFVGKS-KSIMIANISPTISCCEQTLNTLRYSS 329
Db       1492 ILESIKNSKIFRPPNPWTVSILNLLGELHEAQSLKTILIFEIEILFNYFK--INVFDYYN 1549

Qy        330 RVKNFKNKSTCINEED----DTNTERISILDSKGSEMNASSIE--------------NV 370
Db       1550 KCNIIKSRNLPVNSNDLFFRNSFNEHVSVNINSINENNNSMISVNEITPTNYTKSFNTNI 1609

Qy        371 VIKSNHLLSNNNNNKINRGKIN---DKIERNNILKNKSFDKPREGFTSTFGKYSSLNDID 427
Db       1610 IYSEN--LININNDRTNRGGLGLSINKDVRFNMLNNYN---TTQSALATSEATSNNNNIN 1664

Qy        428 KIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNNDSSSMVNNMI 487
Db       1665 N----------NNSSSNNNNNNNNSSSNNNNNNNNNNNNNNNNNNNNNSNDFLSFKGANC 1714

Qy        488 NHMINN-------NINNNINVNNNNNNNNNNNNSH-NNHLPQPNYAFTDTSDFSSLDDMN 539
Db       1715 SLLINNRKEDEGKENNNTRDLNNNNNNININNNAHIMNHLNQ-NHAHLSNGK-NIVNNLI 1772

Qy        540 CH-------LNNNDKSIFLHKKNLR--DNIKLKN----RSSCDNIMNKKKNNLHLARHSV 586
Db       1773 CNSSSDIMLLNNNYRNTFSMNPNLNNVNNKIFKNIDLYKSTNKNILN--TNDVHQKSSSV 1830

Qy        587 GSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHD 646
Db       1831 SKKLT-----------------NPNIMKDN-------HLNTISNYLSN-----DPNINHM 1861

Qy        647 ISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINN-------------------- 685
Db       1862 YTKSGSVNNNNNINNNNNNINNNNNINN-NNNNTFILNNSMDMYQVDNNIYNNMYVNKNM 1920

Qy        686 --SNM--NSNSIYKSNYNSNQSISDVQIRY-------VNEMD----------------- 716
Db       1921 DISNMMNNNENFTKCNNNSNEEMCVKDVKFSNKKKDKLNEIDFKNNIPKSSNNYNINKTK 1980

Qy        717 --TSNKNN---------------------DNIFFDAIS---------------CDNNMY 737
Db       1981 FSVNNKNETSYKLMDMKNGKLVHNKEMIYYENVMKDSYASSSPSMNNINVKGSYMDSSLF 2040

Qy        738 PNITNNNN---------NNNNNNNNNNNIDVENYNNRDGTNNSMKL------------- 775
Db       2041 KNMALNNNLEKLKMNEYNNNNNNNNINNNNSGSNNNQYSNSKNNMMLPPSVVCSPYIVGN 2100

Qy        776 --YAYNSHNLFQ-----PDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYN--LND---- 822
Db       2101 NEMSMKNHMMYQENANNPMNNNNNNNI-NININNNNYNNNININNNSVVFNEILNDEILN 2159

Qy        823 ---KNYLIDLN------------------------------------------------ 830
Db       2160 ILKKNLIISPNIQLIKIHPKFKALIYLAFDSSIKEVITSIVDRFVLIGCITTRELVKKDF 2219

Qy        831 NKEQKDKNIH------------------GC----DNNIIQN------------------- 849
Db       2220 ANESKETIMHKVSLLMATSISSSLVLVSCKKPLKNMLIQNLRNIIEQNFPERNPNLIIVD 2279

Qy        850 ------------------------------------------------------------ 849
Db       2280 DFIKTLINDNINLLYLIIEQVAIEKSSIEIEEIMKPIYASRKYARINKLNINDNTTNRYY 2339

Qy        850 ------------------------------------------------------------ 849
Db       2340 NALPSFLKAHNITLKHIQVYKNFMNLKIFKKLQQKLLSNNNNMEQGNQEYENENAQMNYP 2399

Qy        850 -----------------------RNDFEKKKKTNFYNNNN-------------IVIV-- 870
Db       2400 NKRNMDNNPSNQFDQKVGADFNDYNNNVKRESQIEYHNNNNNNINNGNNQDNGLIKIIKR 2459

Qy        871 -------NNNMGNNNSPRM----------------------KYGLCGSH----TSIDNM 896
Db       2460 DDAPKDDNNNNNNNNNEKTNSKFMDDNKFYHDNNIMDNNEDNNHNLVTSHHKNLSKINTV 2519

Qy        897 KNNEMKNNEMKDNEMKDN---HIKS----NNNNSSSSSSNNNIYNNINDDDTFQNDYCH 949
Db       2520 NNNGLNFIPCKNNSIKENSNFHIGNYNYINSNNNVINFMNNNNLYNSCN----FNYMY-N 2574

Qy        950 NDNTFTIR-RKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNED-- 1006
Db       2575 NKNSFNFNYNKSDTNFNTNRYVNVDNVNTGNNMRMYNSSRNMVDTQVGNMQVIQNNNNSS 2634

Qy       1007 ------IYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKN--NVDN--NVDN-NNVDN--- 1052
Db       2635 GNMDTGIYD------HI-------HDKNNGNSIENSNGKICNIKNISNINNVNSINNRRQ 2681

Qy       1053 -------NVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNS 1105
Db       2682 KKSEYEMNMDNGNNSNIQVGGNINHFEDKTFRGLTEHNDND--SGFNKN-GLFNNNVCNN 2738

Qy       1106 LD--ISSM-YDDTKEILNN----ILLSKYKAEKDN-VIKKYINEDIKNMSLEEIDK---- 1153
Db       2739 MNGYIPNMNYMHNKDVLNNQGNMMNNEQYMMIINNTIMKKGI---IKNNLFNELNNHYIG 2795

Qy       1154 TAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCH-ICNNNPDDQFHFYAYS 1212
Db       2796 TNSNIMNSND--LNKMNAIYKNMISNNENILNNNIIEKLVMNNDMCNGTFTTNSNNISNI 2853

Qy       1213 RLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKLV 1272
Db       2854 NNMNNINNMNNMNNINNMNNNM-----------NNNINNTCINKLMNKDNIMQLDSEIN 2901
```

```
Qy       1273 QDNIKN---------SMDHNN--IHK 1287
Db       2902 NKNIKNNDNIHLNKGAMDNNQKFIHK 2927

RESULT 20
Q8IHV8
ID   Q8IHV8     PRELIMINARY;     PRT;  2275 AA.
AC   Q8IHV8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF11_0418.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014842; AAN36001.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2275 AA;  267960 MW;  FC804F5EB2AED1B7 CRC64;

  Query Match            11.9%;  Score 817.5;  DB 5;  Length 2275;
  Best Local Similarity  24.0%;  Pred. No. 2.9e-25;
  Matches  386;  Conservative  229;  Mismatches  449;  Indels  546;  Gaps   81;

Qy         41 RYKVDMTKYIERHEFIVDKVFDDTVDNFTVY---------ENTIKPLIIDLYENGCVCSC 91
Db         54 RYKLNCKRKKNRK--------DENLDDFEFYIKWKNYDSDENTWEP-----FEN---LSP 97

Qy         92 FAYGQT------------------------GSGKTYTMLGSQPYGQSDTPGIFQYAAGD 126
Db         98 NLKGQARKVALQLTEEIDNCNKNKNEKNDNEGENIEETVVGD---------NINRYIEGN 148

Qy        127 IFTFLNIYDKDNTKGIFISFYEIYCGKLYD---LLQKRKMVAALEN--GKKEVVVKD--- 178
Db        149 I---INNMENNSQQNV-----EQYNNKVYEQTRLVQNTDVLNPNKNKMGFIEIKPNDIDD 200

Qy        179 -----LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILN--------IDLKD 225
Db        201 DDEEIINIMRELEAENM-KKEKNKLRNRNFSLFNMSSNSRLSNEINSYLYENSHKMDVTY 259

Qy        226 INKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANI---------------------- 263
Db        260 IGKKKTGNIKSILSLTNRINNDDNNNNNNNNNIYNNNIYNKRNYNNNYNHCKYSTDLSYD 319

Qy        264 -NRSLLALKECIRAMDSDKNHIPF-RDSELTKVLRDIFVGKSKSIMIANISPTISCCEQT 321
Db        320 GNNNSVHLNDLHHLNDEEKNNLLFLYDSD------DLI---KKDSRIAN--------DIH 362

Qy        322 LNTLRYSSRVKNFKNK--STCINEEDDTNTERISILDSKG---SEMNASSIENVVIKSN- 375
Db        363 YNNLFLSKSIKKLRNKYLESLLN----NNINNKNIFDDESLLKMNLGLSQPSNLCLNNNE 418

Qy        376 --------HLLSN--NNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYS---- 421
Db        419 KSKKCQKPHTFVNTVKNSNKNNENIIND--------FNKEKDESSQNDTSFL--YSTDRT 468

Qy        422 SLNDI-----DKIKKNKKKGLINYK-----------------------STLYNDNTINK 452
Db        469 STNDYYITRKEKIKPLQKESIINNEEDNVQEHHKLLLDKNGNNNNDDDNSTDDNNNDDNK 528

Qy        453 KH----------------NNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNIN 496
Db        529 NNNDNKNNDDNKNNDDNDNNDNNNNNDNNNNNNNNNNNNNNNNNNSNGDNNNNNNNNN 588

Qy        497 NNINVNN---------NNNNNNNNNNSH--NNHLP-------QPNYAF--------TDTS 530
Db        589 NGDNCNNNNDNHNNNDNNNSNNNNNNNHRINNKLDSFEDNNYRENYIFNYLNEKDKTDDH 648

Qy        531 DFSSLDDMNCHLN----NNDKSIFLHKKNL-RDNIKLKNRSSCDNIMNKKKNNLHLARHS 585
Db        649 NSNSYDNSNSNSNMFNYNSNINIDEQNKNIAKNNGDLKNIKR--NVLNKNK--------- 697

Qy        586 VGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESR----NVSNMNGNVLLGLNK 641
Db        698 -GTKF-LPSY-------NTCFKKQICNLRKNMKKKLIVNSMSNKFNISQVNAN-----NE 743

Qy        642 NTH-----------HDISTKDENHNDNKINNGVINIINNSNVN------SINNSNMNSIN 684
Db        744 NNFEEFKQFCLQNGNDQTRKKRLQHEKKKSTNSKNSINNNNNNYYDYNHYFNNKKRSEDD 803

Qy        685 NSNMNSNSIYKSNYNSNQSISDVQIRYV----NE--------MDTSNKNNDNIFFDAISC 732
Db        804 NSFEFNDSYFTNSANTIDQVSKLVQKYLALPSNENIMATDYLLNTTTNNNDN-------- 855

Qy        733 DNNMYPNITN--NNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNK 790
Db        856 -NKNNDNINNEDNNKNNDNINNDDNNNNDDDNNNNDDDNNNN-------------DDNNK 901

Qy        791 NTSNIQNINTNKNNQDGNV-------------------------NYSMNFCHYNLNDKNY 825
Db        902 N-----NDDNNKNNDDKNICKKGRKTNNIKKRKILVNDNNISLGNYSNNE-YYDYHDVN- 954

Qy        826 LIDLNNKEQKDKN-IHGCD-NNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMK 883
Db        955 ----NNSNSNIRNMIHNKSLSDFIDEDKIFKQKSGKNIVKN---LFCNKKSYSNNSAYVN 1007

Qy        884 YGL-------CGSHTSIDNMKNN-------EMKN---NEMKDNEMKDN----------HI 916
Db       1008 RGISEYAEENIGDKNFINNQKCNNYFLNLKELQNIYASEKYGKSKKGNKNLVSKKINDYI 1067

Qy        917 KSNN--NNSSSSSSSNNNIYNNINDD-------------------DTFQNDYCHN-DNTFT 955
Db       1068 YNNNSGNNSSNNNSSNNNSSNNRNNNIRKSYINKLTESEYDSVYETLINRYCKNTTNKYS 1127

Qy        956 IRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNK---EM 1012
Db       1128 KKINNDENGNTNEPINDIKYY----INDNNNNSTSNSRPSLIHDNLSEND--FNKIGDEL 1181

Qy       1013 EGKHIRL-DDQDKYDDNDNNNVDNN-NKNNVDNNVDNNNVDNNVD--------NNDKNNV 1062
Db       1182 KRKQSQITKENQKYDAHVDDDIDDDEEEENTDKN--NNTIRSSIQSDIIYKILNDEKSKS 1239

Qy       1063 DNNNV--------------------DNDDDDVDF------------------HNIKNFNNNEYLS 1089
Db       1240 QEQNVHILNKDNINEFSEFKKKDGENDEMVFFSSDNMKSNIYVAKKNYSSESFNNKHIMK 1299

Qy       1090 YFQ----------------KNVD--------TIINNCLNSLD-ISSMYDDTKEILN---- 1120
Db       1300 KTQHLYNMDEIKNKNKNKSKNVSDSTNENNHTNHNNNINKYDGLNSMTINIKDILDTKND 1359

Qy       1121 ------NILLSKYKAEKDNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFK 1174
Db       1360 KRDYYINVLENVLSALKEN--RTYNSKDL-NISKE--NENANGHIRSNSYTDTSSILLNR 1414

Qy       1175 KNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENL 1234
Db       1415 KIPEISGEKENYENLRNLKRRTFHNDSIE-------YSRTPK------------------ 1449

Qy       1235 RLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKLVQDNIKNSMDHNN 1284
```

```
                |:|::     ::|||                    | || |: | : :  | ||
Db       1450 RVLFEINKEDHQNKK----------------LKNKNLIDDKVGDDDDDNN 1483

RESULT 21
Q8IBK6
ID   Q8IBK6      PRELIMINARY;      PRT;  2041 AA.
AC   Q8IBK6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   MAL7P1.132.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA   Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844506; CAD50994.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2041 AA;  243553 MW;  B5BDD4D494E9A4C6 CRC64;

  Query Match            11.9%;  Score 816;  DB 5;  Length 2041;
  Best Local Similarity  24.3%;  Pred. No. 3e-25;
  Matches  332;  Conservative  201;  Mismatches  438;  Indels  398;  Gaps   62;

Qy        199 LLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQT 258
              :|  |  || |: :: :    ||: : | |    |   :     |       :|||
Db        380 ILNDILRNSNNNNNNNYNN--NINKLNNNNNNCYNKTKLLTYNTSS------LRNKQKDD 431

Qy        259 DGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCC 318
                  :  |  ||  |     : | :| | |        : :| |    : ||: ||    :
Db        432 EEKKREKKLNLLKNIIVNNPNKKCNISFNDDFYDSLAQDNFYVNTSSIVDYNIGTIPN-- 489

Qy        319 EQTLNTLRYSSRVKNFKNK--STCINEEDDT---NTERISILDSKGSEMNASS------- 366
              :   |          ||:||       :|: |        |    :| |    | |  :
Db        490 DDVFNK-------KNYKNNHIKDHMNDHDKNMHKTKEHNNIYDKNIYEDNEKNGYDKNKM 542

Qy        367 ------IENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKY 420
                    | : | |      | :|     : :: | ||:   |::|    :   |: |
Db        543 YNFLYDDLNTIDKMNETPYIMNKHKFYNENVPNE-EDNNMTIEKNYD--YFSYDKTYKKM 599

Qy        421 SSLNDIDKIKKNKKKG-LINYKSTLYNDNTI--------------NKKHNNN-------- 457
              :: ||||  :|   :|  | | |:|     :|:   |              |    |||
Db        600 NNFNDIDFVKNKEKNGNLLNNNINMYDNVKIGLNQGVDIFRTPMRNYNVNNNYEEDFYPY 659

Qy        458 -NNNN-----------------NDNNDNNNDNNNNNNNDSSSMVNNMINHMI-------- 491
                ||:|                | ||:|||:|||||||     :    :       :
Db        660 KNNSNISVNKNEKQVFQYNSFENSNNNNNNNNNNNNNNDCKNEAHKYDKHIYDKYNFNNL 719

Qy        492 ----------------------------------NNNINNNINVNNNNNNNNNNNNSHNNH 518
                                                 ||| ||| | |||||| |||||::||:
Db        720 GYKTNDKGHDHYDENLSKMSPFHKFNDNNNNNNNNNNNNNNYNNNNNNNYNNNNNNNYNNN 779

Qy        519 LPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKK---NLRDNIKLKNRSSCDNIMNKK 575
                  || :           | ::| |:  |    |    |    | |||  :   : |:|
Db        780 ---NNYNYN-----------NNYINENNLKHFEQNKIGANNGRNSFLKNSNFLFDDMDKY 825

Qy        576 KNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNI---NKMEDNTPKDILY-ESRNVSNM 631
              ||: :            |:   :|     : :  :|:||   | :: |   ::|| :: |  |:
Db        826 KNDQY------DKINK-NYHSVVERKSYLYKNNIPINNTLKKNDQNELLYLDNPNEINI 877

Qy        632 NGNVLLGLNKNTHHDISTKDEN-----------HNDNKINNGVINIINN-SNVNSINNSN 679
                  |    |  :    |    |                :||:|:|   :  |  ||   :| | |
Db        878 FRNTDKGNDHFTQDTFLKKHTNIIEMNKILNKYNNDDKLNT-LDNYDNNIFYMNKIKNTN 936

Qy        680 MNSINNSNMNSNSIYKSNYNSNQSIS-----------DVQIRYVNEMDTSNKN------N 722
                |: ||:|  |:::|| :|||:| : :                 :  |    |    |||     |
Db        937 NNNNNNNNDNNDNIYNNNYNNNYNNNFYMRSEKSPHCNNQDDIYNYRDIFNKNKWDFLKN 996

Qy        723 DNIF----------FDAISCDN---NMYPNITNNNNNNNNNNNNNNNIDVENYNNRDGT 769
              |::                   :   |  |      |  |    |||||:|||||||||::    :|   :
Db        997 DDMLLNNNMTSLNNMEKNKCVNNTINKYTPYINNNNNDNNNNNNNNNMNSNIFNIHNDK 1056

Qy        770 NNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVN--YSMNFCHYNLNDK---N 824
              || :              :  | :|:  |              ||| | |::         |  |  :   |
Db       1057 NNMVP---------YSDDKDKDKIN----KIFKNNYDENIDDIKLSNICFNNKKENDIYN 1103

Qy        825 YLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKK-----------TNFYNNNNIVIVNN 872
                : : |:       |::   | ::: :| |   :                   | |||      ::|
Db       1104 VIKNYNSPMHISTNMNNAHNTVLKKKNFFNLENNNVPPIDHITYNHNNLYNN-----IHN 1158

Qy        873 NMGNNNSPRM--------KYGLCGSHTSIDNMKNNEM-----KNNEMKDNEMKDNHIKSN 919
              |: || |   :          |  :  ::|:    :  || :         |::|:  |    |  :|| |
Db       1159 NVHNNKSETLNLFGRTNNKSDILNNNTNNSYIYNNNILYKGKNNDDMEYKETDDIYIKDN 1218

Qy        920 N----------------------NNSSSSSSSNNNIYNNINDDDTFQN---------DY 947
              |                           :| :         :||  |::||:  :|
Db       1219 NISQKYYPNKINIFHKDDEYEKLQDNVTKLKKDYSNI--NVHDDEYLKNMNTKYNLHKQN 1276

Qy        948 CHNDNTFTIRRKNNTNI---NSNIYQN-----DDI----IYTINSLND----------YM 985
                |:|   :  | ||||    ||  |:  |            :||      :|: : :|:     |
Db       1277 IFNNNEMDLLEKKNTNIYFPNCNIFNNFNKQENDIKNKNLYSYHDINNKNKERFEYIKYG 1336

Qy        986 SNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDND---NNNVDNNNKNNVD 1042
                | :          :|       |:|   ||       :| |:      | : ||     | :: ||| |:
Db       1337 HNHIHQHANNKNFPLTHFNDDDQNKNY---NIFLNTPTKGNINDIGYNKHISNNNYNHFK 1393

Qy       1043 N-----NVDNNNVDNNVDNNDKNNVDNNNVDNDD--DDVDFHNIKNFN---NNEYLS--- 1089
              |       :  ||: ::| :||   |     ||:   |:        :  |           :|    || :|:
Db       1394 NVHKKGDTQNNHNNHNNNNNSNNFFTNNHQGNNKALPPLPFKTNNKYNGSQNNSHLTKLK 1453

Qy       1090 YFQKNVDTIINNCLNSLDI------SSMYDDTK----EILNNILLSKYKAEKDNVIKKYI 1139
              :|: |   |   |: :   :           :::| :|          :|||:          :|    ||
Db       1454 FFEHNDKTHKNDSIQETNQMINGGGNNIYSNTNISNINMLNNM--------NENNHSLYI 1505

Qy       1140 NEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICN 1199
              || |  |:  :|  :             ||:||    |       ||   ::  :: | |  :|
Db       1506 NE-IVNIKKDEHNN-------------TNLLVMNHHN-----NNNNNNNNNNMVMMMSCN 1546

Qy       1200 --NNPD----DQFHFYAYSRLEKDIINLIMLRQIWCESENLR---LLYQ-------FLVV 1243
                :||       :| | |                          | |:|    :||         |||
Db       1547 DTSNPQKMKLNNFPF---------------------SHNIRNKNILYNEDKDQELFLVE 1584

Qy       1244 EYQNKSANSVLLNVSSN-----NGDIILLNKKLVQDNIKNSMDH---NN 1284
                : : |  : :  | :  |         |  |  |   |  |   :|  |::::  :     ||
Db       1585 NKKKRDAENYMFNETYNALKNTNNDNIFHNNKRKKDCIESTLTNLYTNN 1633

RESULT 22
Q8IAU8
ID   Q8IAU8      PRELIMINARY;      PRT;  1774 AA.
AC   Q8IAU8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   MAL8P1.104.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA   Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844507; CAD51262.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1774 AA;  200169 MW;  182115F3C2F4A8A4 CRC64;

  Query Match            11.9%;  Score 815.5;  DB 5;  Length 1774;
  Best Local Similarity  24.5%;  Pred. No. 2.8e-25;
  Matches  318;  Conservative  199;  Mismatches  411;  Indels  369;  Gaps   55;
```

```
Qy     212 SSRSHAILNIDLKDIN---KNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLL 268
Db     586 SNANNGLLNLNNTNLNNSINNTINGSLILGGMNNNGHTNNNINNNHNNNGNAINMNMNGM 645

Qy     269 ALKEC-------------------IRAMDSDKNHIPFRDSELTKVLRDI----------- 298
Db     646 GIMNSSMNISGLSTNLNMNNLYGDIRSLGGSTSNINLNNINSMNELNNLKSGSISNISNC 705

Qy     299 ---FVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISIL 355
Db     706 DENMIGLNKNIYSTNNGPLIG---NMINNNVYNANI-NYNNYN--YNSINTNNSSNNSFL 759

Qy     356 DSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNI--LKNKSFDKPREGF 413
Db     760 NS-------------VNYNTAYAHGNNPSISPGG-NNNMSTNNYYSIQNK--------- 795

Qy     414 TSTFGKYSS-LNDIDKIKKN--------------KKKGLINYKST----LYNDNTINKK 453
Db     796 YSNFGKYSSNLNYVNPLSQSINEVKSMAGEDYCDIDKRSIDNLGSSKNSFANNKNLLGKI 855

Qy     454 HNNNNNN---------------NNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNN 498
Db     856 NNNNNNNIISSSSSTSHHINFKNTISNNNSNINNNNLNINSTTSVHPMNNP--NMNDTNN 913

Qy     499 INVNNNNNN---------NNNNNNSHNNHLPQPNYAFTDTS-----DFSSLDDMNCHLNN 544
Db     914 LNIKRNMKNESFISDLSTNKNELNNMGDHVGMENLNMLVQNMNIKRDEKYKDEKINNNNN 973

Qy     545 NDKSIFLHKKNLRDNIKLKNRSSCDNIMNKK----KNNLHLARHSVGSKLTMFSYDPQKN 600
Db     974 MSKNMML---NLNNNSSISNVNNLNNMMNSNNDLIKNNNTTYMKNINMSISNINNGVTNT 1030

Qy     601 KDNTFFK--SNINKMEDNTPKDILYESRNV---SNMNGNVLLGLNKNTHHDISTKDENH- 654
Db    1031 VLNTDLKVSSNLNQNGINVTNSTTYISANSLDNNNNNNNLNSGNNINDHGNNSGKGINHT 1090

Qy     655 ------NDNKIN---------------NGVINIINNSNVNS-----------------I 675
Db    1091 NPTIINNDNKINVASHNKGNTSIDGNWINGPINNTNSSNKYSEGKDKDANTEDNKIGYKA 1150

Qy     676 NNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKN-NDNIFFDAISCDN 734
Db    1151 NDNNNNSSNNNRNNNGKVLSNNKNSNKG---------NNKNSKNKSKNNNDPNKKASGDN 1201

Qy     735 NMYP-----NITNNNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNN 789
Db    1202 LNEPDYKENNMNNNNNNNNNNNNNNNNNN---NNNNNNNNNNN------NSNNNNSNNNN 1252

Qy     790 KNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQN 849
Db    1253 SNNNNNSNNNNNNNNNNNNSNNNNN----NSNNNN-----NNNNNNNNNNNNSNNNNYNN 1303

Qy     850 RNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDN 909
Db    1304 SNNNNNNHSTLISNNDNTIVSNNDNEDNS----------NHTTNNNEQDKDDKNNNMK-- 1351

Qy     910 EMKDNHIK----------SNNNNSSSSSSSNNNIYN-NINDDDTFQN-----DYC----- 948
Db    1352 -TEGNHSKFSKQHGNMNTRNKNNDKNNFVNNSNLVNSNVEDVNNFNNIINQRNHCNKNYI 1410

Qy     949 -----------------------HNDNTFTIRRKNNT-----NINSNIYQNDDI-IYTI 978
Db    1411 NFIKEDQQEVDQKELHASNISHMLNEDNSNEIMDKNTSYNNPYNNNANYYKSDNIKLYR- 1469

Qy     979 NSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEM-EGKHIRLDDQDKYDDNDNNNVDNNN 1037
Db    1470 ------------GFEKNSSHQLILKNNFSNNSNMTENEGIIMHDMNKMNNPNNADILLGN 1517

Qy    1038 KN-----NVDN----NVDNNNVDNNVDNNDKNNVDN-NNVDNDDDDVDFHNIKNFNNNEY 1087
Db    1518 ENLLGIKNLNNYMNPNVHNTERNNELNNNNTNMITNKKNVNSCNNDMLHKQINNFN---- 1573

Qy    1088 LSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMS 1147
Db    1574 ----IININKVHNNISNSL-----------FSNN--MSNYNVVSENV-------NINNNS 1609

Qy    1148 LEEIDKTAQSIYEKRK-----------VLLTKLLL------LFKKNVDTQINNETSDLRK 1190
Db    1610 --NTNNTANNNFMNKKNSKANLNVFNQSLINNLNLNLNNFYLNNENSFNKINMNYSNINS 1667

Qy    1191 DLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSA 1250
Db    1668 TSYYFNTNDNNMNNTQHYYSSFIPEHFATDKIVTKQ---SSEN----------------- 1707

Qy    1251 NSVLLNVSSNNGDIILLNKKLVQDNI----KNSMDHN 1283
Db    1708 -----NNNINNSSTMYKEKK-EEPNLNFKEKNFVDKN 1738
```

RESULT 23
Q8IDR0

```
ID   Q8IDR0      PRELIMINARY;      PRT;  3848 AA.
AC   Q8IDR0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF13_0235.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844509; CAD52559.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   3848 AA;  454964 MW;  57D2700BEA3F5A9A CRC64;

  Query Match             11.9%;  Score 813;  DB 5;  Length 3848;
  Best Local Similarity   24.0%;  Pred. No. 7.2e-25;
  Matches  330;  Conservative  241;  Mismatches  435;  Indels  368;  Gaps   66;

Qy     126 DIFTFLNIYDK------DNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDL 179
Db    1288 DIHNY--IYEENPFNILDNIK--------------YELKDVYNQMMNLKKTDNILTIND- 1330

Qy     180 KILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
Db    1331 KLLHNLLKLDNSTHQMD---IPNISYDNKWKGSSNNH---NNDLNNNNNNY--------- 1375

Qy     240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSD-KNHIPFRDSEL---TKVL 295
Db    1376 -------YHNYYLNNNNFNENKNVLEELLSNNNCFNRINNQFKNKDNFKEPNLMTHNDIA 1428

Qy     296 RDIFVGKSKSI--MIANISPTISCCEQTL--------NTLRYSSRV-KNFKNKSTCINEE 344
Db    1429 DDISICSNEDVNDIIYNINRTDDILKRLMYLSNNYYNNMIEYDNYILHSYQHNDHILNEL 1488

Qy     345 DDTNTERISILDSKGSEMN-----ASSIENVVIKSNHLL-------SNNNNNKI------ 386
Db    1489 NNNTVTYYNNIKKKKNFNNQLLPRTKRNEDFIFNNNHIVTPNNIYNNNNNNNKLADSEIM 1548

Qy     387 --------------NRGKINDKI-----ERNNILKNKSFDKPREGFTSTFGKYSSLNDID 427
Db    1549 ENVMQTCNQNNYQFNRTESNDNLKNIPQENNFLLPNSNMEK--------FGELINV-DMN 1599

Qy     428 KIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNN------NDNNNNNNNDSSS 481
Db    1600 NIENN-----INMFNNMYNN--VNNNINGNNINGNNINGNNIIGNNINGNNINNNNINSN 1652

Qy     482 MVNNMINHMINNNIN-NNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 540
```

```
Db       1653 NINS--NNINSNNINSNNINSNNINSNNINSNNINSNNINSNNI---------NSNNINS 1701

Qy        541 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600
              :: ||      ::   | ::|:   | :  :|| |:        :||       : ||   |
Db       1702 NIYNN-----MYNMNYKNNVISSNNNYINNINNE--------QHSNPMNYNTYGYDLSNN 1748

Qy        601 KDNTFFKSNI----------NKMEDNTPKDI---LYESRNVSNMNGN------------- 634
               :: |: :||          | :::   | |   :|:| :: |:  |
Db       1749 VNHLFYNNNIMDGRYNISNNNIIQNENIKGIYPHIYKSNHMVNVETNNFDIYDNVPNKYT 1808

Qy        635 ---------VLLGLNKNTHHDIS-TKDENH-NDNKINNGV--INIINNSNVNSINNSNMN 681
                       :   :  ::||:   |||: | | ||| :  :|: :|::::| ||:| :
Db       1809 SMDNKDFVGEYISDREERYYDINILNDENNINKNIINNNINDMNVYDNNSIHSNNNNNFD 1868

Qy        682 SINNSNMNSNSIYKSNYNSNQS-------ISDVQIRYVNEMDTSNKNNDNIFFDAISCDN 734
               | | : |:|:|     ::|  |          :::  . | || :||:          | |:
Db       1869 YIKNIDNNNNNINDRVFDSRNSYECSFPLLTNGPYYYENENNTSSN---------ILCNQ 1919

Qy        735 NMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSN 794
              ::  |:||||  |  :| :||  |  : :  ||      | :       :|:: |
Db       1920 DV--NLTNNNFINYESNISNNYNQLISDNNNEGNINMNKNFNVFNNNVL----------- 1966

Qy        795 IQNINTNKNNQDGNVNYSMNFCHYNLNDKYNLID--LNNKEQKDKNIHGCDNNIIQNRND 852
              | |   | || |  |       |    |: | | :|| :|      : |:: :||||:|   :
Db       1967 INNCTINDNNVDNNKMIDSNMIDGNMIDSN-MIDSNMNGNNMNENNMN--ENNIIENNMN 2023

Qy        853 FEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMK 912
                 : |    ||    |: |||  ||           :  :: : :|: :| |:| : || |
Db       2024 GNNIIENNMNGNN---IIENNMNGNN-------IIENNMNGNNIIDNNMIDNNLIDNNMI 2073

Qy        913 DNHIKSNN---NNSSSSSSSNNNIYNNIN---------DDDTFQND-----YCHNDNTFT 955
              ||::  ||     ||    ::  :||:|::||          | ::|:|           | : |
Db       2074 DNNLIDNNLIDNNMIDNNMMDNNMYDHINNFTINSPMIDMNSFEEDPKNYNMIHINETDI 2133

Qy        956 IRRKNNTNINS--NIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTN----EDIYN 1009
               |    |  ||:  | |::|:   |  : |  :| :           |        ||   : ||
Db       2134 INLTNEDLINNSENYYKHDN---TSSYLLNYQNCKNFELKNVSNVEIEQTNIMDSNNTYN 2190

Qy       1010 K-----EMEGKHIRLDDQ-DKYDD--------ND------------------NNNVDNNN 1037
              |     |     || |::: :| :|        ||                  ||: :||
Db       2191 KITEENETNSSHILLEEKINKTNDMCKNDTEINDHSLAILLKKNPSEKKKIMNNHHNNNK 2250

Qy       1038 KNNVDNNVDNNNVDNN-----------VDNNDKNNVDNNNVDNDDDDVDFHNIKNFN--- 1083
              |||     |: :  ||          :|:|| |:  |||    ||  |:      :||||  |
Db       2251 KNNKKKKKDDLHEHNNKEKLQNCHNEYLDSNDYNS-DNNT--NKDNVFVINNIKNTNCNI 2307

Qy       1084 --NNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKD-NVIKKYIN 1140
                ||: :|    ||||      |   :    |||          ::|| |  :|   :|  |
Db       2308 NQNNDIIS---KNVD----NSSKNNDISNEKDN---------INKMKLKKMLEITNKLIG 2351

Qy       1141 EDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN 1200
               :  :  :  :         |   |    |    ||           |   : :|   :   : |
Db       2352 KKYRGICYDPTRNGWSTFVYKDGVRYKKFFSSFK------YGNLLA--KKKCIEWRLKNL 2403

Qy       1201 NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSN 1260
               ||    | |::|   |:   | ::        ::|     |: : : | ||  ::  |: :|
Db       2404 NPSS--HAYSFSLKAKEEFNEVL-------NDN----YKDVGILYDNKKDSNNNNNIKNN 2450

Qy       1261 -------NGDIILLNKKL-------------------VQDNIKNSMDHNNIHKK 1288
                      | ||: :|  |                  |:|| ||: :    |  |
Db       2451 GNQDDDGNRDILYVNAFLNIYNDENVKEKCFTDDNDNVKDNEKNTKEIKIIKNK 2504
```

RESULT 24
Q8T2M5

```
ID   Q8T2M5      PRELIMINARY;      PRT;   2472 AA.
AC   Q8T2M5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 278.5 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115593; AAL92295.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2472 AA;  278497 MW;  30CCF7157D4008A7 CRC64;

  Query Match             11.8%;  Score 810.5;  DB 5;  Length 2472;
  Best Local Similarity   22.3%;  Pred. No. 6e-25;
  Matches  302;  Conservative  191;  Mismatches  400;  Indels  459;  Gaps   41;

Qy        214 RSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKEC 273
              :| ::|| |  : :  |                :   :   :   |     :  | ||:   |:
Db        291 QSLSLLNFDSDEGDNNNF----------NNNSNNNNNNNNNNNNNNNNNRNDDNLR-V 338

Qy        274 IRAMDSDKNHIPFRDS-ELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVK 332
              | :| | :      || : ::    : : ::|    |:|  :|      |:: :||
Db        339 IEELDEDDDDYLSSDSYDSSRPSSSVDLNQTKMFTTADIISSI-----PLSSSSFSS--- 390

Qy        333 NFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKIN 392
                          |  |  | ::|  |::  :| |:   :|   :||||:  |    |
Db        391 -------------TTTFESNSNINSSNSKIKLPTI-NLSQMNNINQNNNNNSNNNNSNSN 436

Qy        393 DKIERNNILKNKSFDKPREGFTSTFG-KYSSLNDIDKIKKNKKKGLINYKSTLYNDNTIN 451
                       |  : | |            ||     |||: : :       |   ::  |:|:|:
Db        437 SNTNNGNNTQPKKF---------TFSLPLSMLNNNNNVNNGN-----NSNNSNNNNNSIS 482

Qy        452 KKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNN 511
               :|||||||:||:|||:||||||||::::  ||  |:   |||  |||  |  |||||||||
Db        483 GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 542

Qy        512 NNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNI 571
              ||::||:                  | : |||: :                || :
Db        543 NNNNNNN------------------NNNNNNNNN---------------GTSSSSSA 567

Qy        572 MNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTF--------FKSNINKMEDN-TPKDIL 622
                 : : |:   :  :||   |     |                :||| :    : | :
Db        568 TAASASGILLSPRLMSPRLTTVIQAPASTASNPLTVMSPRRQSMANINSLAPGWSSKSLE 627

Qy        623 Y---------ESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVN 673
              :          :|:: |    :  :  |    |::  : |:|:|  ||  ||   ||:| |
Db        628 FLHSDMKKSKKSKSDSQQERDDTVQLGSIIEFDVNNNNNNNNNNNNNNNNNNNNNNNNNN 687

Qy        674 SINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCD 733
               : ||:|  |: ||:| |:|  |:    : |||  | |                  :| ||:||
Db        688 NNNNNNNNNNNNNNNNNNNNNNNNNNNSNNSSSN-----------NNINNNNI-------- 728

Qy        734 NNMY-----PNITNNNNNNNNNNNNNNNNIDV--------ENYNNRDGTNN-SMKLYAYN 779
              || |         :| ||: ||||||||||   :          |||: : :|| : ::|
Db        729 NNNYNSNYNKSINNNSINNNNNNNNNNNSSINNNGINNGSNYNSNNNSNNINNQIYQQQ 788

Qy        780 SHNLFQ---------------------------------PDNNKNTSNIQNIN--TNK 802
                 |:                                     |  |  | :||  |  |    ||
Db        789 QQQQFKQQQQIQDQVYKQQDESITPRELSRRNSSSGLILNSPSRNNNINNINNNNGSTKK 848

Qy        803 NNQD------GNVNYSMNFCHYNLNDKYNLIDLNNKEQKDKNIHGCDNNIIQNRN----- 851
                           |:| | |    : |:|: | : ::||    :  | :  :|| | |::
Db        849 GTTKLPTLKLQNINNSYNNNNNNINNNNNNINNINNNNINNNNNNNNNNNNINNQSVPLIP 908

Qy        852 ----------------DFEKKKKT----------------------------------- 859
                               :|: ||
Db        909 NTVNNIPNAPPPPVYEPVDKQYKTIKHPLTRVSSEPSIRFNGMPIQQLQLQQLQQPQIQQ 968

Qy        860 ------------------------------------------------------------ 859

Db        969 QQQQQPQIQQQQQQQQQPQQQQQQQQQQPQIQQQQQQPNVNLNNNGKPYIPKVQLNTLKT 1028
```

```
Qy        860 --------------------NFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKN 898
Db       1029 QDPDKPLFSPRLQSQMSPHPSNMYNNNN----NNNNNNNNN-----------------NN 1067

Qy        899 NEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDN-TFTIR 957
Db       1068 NNNNNNNNNNNNNNNNNNNNNNNNNNNSNNYNSNNNYNNNSNNSYSYNNNNTTNNNSSININ 1127

Qy        958 RKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLH-------------------------- 991
Db       1128 NNNNSNNNFNYNNNNNNNNNNNNNNNNNNNNYNHSNGGNNSNSSLNEQSLDQSFTSPMIF 1187

Qy        992 -----------FKEKYTYPTLSTNED---IYNKEMEGKHIRLDDQDKYDDNDNNNVDNNN 1037
Db       1188 EIANKLKAMDPSKKKLIMDLLSSDDSSTTIYQHLMQQQQQQQQQHNQYDDGNDNNNSNSN 1247

Qy       1038 KNNVDNNVDNNNVDNN--VDNNDKNNVDNNN----------VDNDDDDVDFHNIKNFNNN 1085
Db       1248 SNNNNNNNSNNNGNNNKPVKKISPYNTPHSSPMMSPSPSPRIDLRDFNLDYSNIE----K 1303

Qy       1086 EYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKN 1145
Db       1304 KRPSSIRKKIET-----------QQHHQQQTNLQPQQILLQPPPPSPSLTPRSISTPILS 1352

Qy       1146 MSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQ 1205
Db       1353 SQI-NVNNVQSSTSNRPK---------FSLNFDQLGNN---------------NNNVNTP 1387

Qy       1206 FHFYAYS-RLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSAN--SVLLN------ 1256
Db       1388 LHTGASSARLSSN------------NNNNNR---KIPTGGYPNNTTSPASVLVSPRRFSI 1432

Qy       1257 ----VSSNNGDIILLNKKLVQDNIKNSMDHNN 1284
Db       1433 SPPTINNNNNNINNNNNNNNNNNNNNNNNNNN 1464
```

RESULT 25
O97277

```
ID   O97277      PRELIMINARY;      PRT;  1200 AA.
AC   O97277;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Kinesin-related protein.
GN   PFC0860W, MAL3P7.1.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=99376085; PubMed=10448855;
RA   Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA   Churcher C.M., Craig A., Davies R.M, Devlin K., Feltwell T.,
RA   Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA   Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA   Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA   Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA   Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT   "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT   falciparum .";
RL   Nature 400:532-538(1999).
DR   EMBL; AL034559; CAB39023.1; -.
DR   HSSP; P33176; 1BG2.
DR   InterPro; IPR001451; Hexapep_transf.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00132; hexapep; 4.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   ATP-binding; Coiled coil; Microtubules; Motor protein.
SQ   SEQUENCE   1200 AA;  138263 MW;  7673746EFB604614 CRC64;

  Query Match             11.8%;  Score 810;  DB 5;  Length 1200;
  Best Local Similarity   25.1%;  Pred. No. 3.2e-25;
  Matches  309;  Conservative  201;  Mismatches  370;  Indels  352;  Gaps   50;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPR----YKVDMTKYIERHEFIVD 58
Db        256 SNVKVAVRIKPIGESEE-----NIVSIFNKNYVLIEKENEKECYLLSQKK--KQSTYVFD 308

Qy         59 KVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPG 118
Db        309 SVFDVNATQEEVFFQTAKPLIPHVFK-GINCTVFAYGATGSGKTYTMLDDK-----NQNG 362

Qy        119 IFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRK-MVAALENGKKEVVVK 177
Db        363 IVQLSLLELFTIINEKKCRNIK-VLMSFLEVYNETIRDLLGKEKNKTLEVQEDVAEVKVS 421

Qy        178 DLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNI----DLKDINKNT--S 231
Db        422 NLCEIEVNNYEQAMLLINEGVKNRKMSPTRANKVSSRSHAILQIYVYNEILDDNMNTISY 481

Qy        232 LGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKN----HIPFR 287
Db        482 KAKLCFVDLAGSERASATSNKGERFK-EGSYINQSLLALANCINSLASNRNISKVRVKYR 540

Qy        288 DSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDT 347
Db        541 DSKLTHLLKNSLEGNCLVVMIANINPSRTSFQESNNTLKYAFRARNIK---LCATVQTND 597

Qy        348 NTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFD 407
Db        598 NKE--------------SDIEKILKKNENL------------------------QKEYD 618

Qy        408 KPREGFTSTFGKYSSLND----IDKIKKNKKKGLINYK-----------------STLYN 446
Db        619 -------TLLGKYTNLKEFFFIINVINQLYKKQISCYKLIENISDNMSSMELKQDITMY- 670

Qy        447 DNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNN 506
Db        671 DQLVKMK--------SDEYRKKVDSLKDLYQEEKQFLNNLFDTFLEKNLNYVINSKDVND 722

Qy        507 NNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRS 566
Db        723 NNK-----------------------SLLEEM---------IFFKHNENKVNENFLVNEK 750

Qy        567 SCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESR 626
Db        751 VVD------KNNV---------------------------LNGNVMVDEN----VVDKN 772

Qy        627 NVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNV----NSINNSNMNS 682
Db        773 NV--LNGNVMVDENVMVDENVMV-DENVMVDE------NVMVDENVMVDENVMVDENVMV 823

Qy        683 INNSNMNSNSIYKSNYNSNQSI-----SDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMY 737
Db        824 DENVMVNENVVVDKNVSVNENIVVSEKHKVGLSAEGKSESHNKNNKD---DIEDNDKDTI 880

Qy        738 PNITNNNNNNNNNNN-NNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQ 796
Db        881 KDIHNNNNSSDNNDDEYQSANSPVES-------------------DIVKKEKKKKIP--I 919

Qy        797 NINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKK 856
Db        920 NMETKKKRTMNGTKDPIHKTPYDIN----IVGILNKEDVSNKSNDYNTNKNIEKNNYEKK 975

Qy        857 KKTN-FYNN-----NNIV--IVNNNMGNN-NSPRMKYGLCGSHTSIDNMKNNEMKNNEMK 907
Db        976 GEYNPFHNNLTDMQNSILYNIINNNVENSPHSPRMK------------------------ 1011

Qy        908 DNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSN 967
```

```
Db     1012 ------------KNVAKMLLKGNLNTANFILDDDTIK------------------NMNSN 1041

Qy      968 IYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDD 1027

Db     1042 ---------KISDKHNMKSNNILN------------NEN-------GK---INDKSKKCK 1070

Qy     1028 NDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEY 1087

Db     1071 NINNN-NNNNNNNNNNNNNNNNNNNSSSSSSGKVDGINI-----------LNNSNTNER 1118

Qy     1088 LSYF----QKNVDTIINNCLNSLDIS------SMYDDTKEILNNILLSKYKAEKDNVIKK 1137

Db     1119 LHTFSGVYSLNLNDEIKIEINKKDMEKNDIHLTSIDTISKIHARDLLKENKRKLEN---- 1174

Qy     1138 YINEDIKNMSLEEIDKTAQSIYEKRKVLLTKL 1169

Db     1175 -FQENIKHEHKDEV-----SLYVKKKKIKKKM 1200
```

RESULT 26
Q8I5X5

```
ID   Q8I5X5      PRELIMINARY;      PRT;  2612 AA.
AC   Q8I5X5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFL0350C.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014845; AAN36159.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2612 AA;  311421 MW;  C950B6E466E3D1C5 CRC64;

  Query Match            11.8%;  Score 808;  DB 5;  Length 2612;
  Best Local Similarity  24.2%;  Pred. No. 8e-25;
  Matches  373;  Conservative  220;  Mismatches  434;  Indels  514;  Gaps   78;

Qy        2 NSKIKVVV---RKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDM--------TKYI 50

Db     1174 NEKLKDEIEKERRNMIQNLEEEKKE---------FKLYL-EQKYKENFENEKSGLAKKFD 1223

Qy       51 ERHEFIVDKVFDD------TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYT 104

Db     1224 EENEKLQNEIGNEKRKLHKERDNFEEQKK--------IYEEEFRNKCEKY-EESIQKKYD 1274

Qy      105 MLGSQPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYE--------IYCG---- 152

Db     1275 VLEE-------------------------DKNKMKYLIIKEQEELENYKKNIYLDIEEE 1308

Qy      153 --KLYD------------LLQKRKMVAALENGK--KEVVVKDLKILRV--------LTKE 188

Db     1309 KEKLYVQQEKLNLEKENLLVEKEQIEIELKNYKNFKEKEENDIKIKIINLSQQQEDLNKE 1368

Qy      189 -ELILKMIDGVLLRKI-------GVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240

Db     1369 KENIEKEKEQIEKRKYDLDEREESLNNDKAQMEESRKIFDEQLDKIKKNKE--ELLNYDK 1426

Qy      241 AGSERGADTVSQNK-------QTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTK 293

Db     1427 ELKNREIHLMEKEKEYNKKLEESNKKKESLNSFDVELKEYSSKLQEREKKLKEKKNELQK 1486

Qy      294 VLRDIFVG-----KSKSIMIANISPTIS--CCEQTL-----NTLRYSSRV---------- 331

Db     1487 V-KDQLVNYKNSLKQKEMQLQNIEKREKELLDEQTVIQIDRNSLEAEKKQFILTKEKHIK 1545

Qy      332 -------------KNFKNKSTCINEEDDTNTERISIL---------DSKGSEMNASSIEN 369

Db     1546 DTEFIQEQLKLLHEQLKNKEKSLKEKEN----EINLLSKLQDCRNKSGKGAIISLKHLEN 1601

Qy      370 VVIKSNHLLSNNNNNKINRG-----------------------------KINDKIER-- 397

Db     1602 RKVSTN--LSRTSNAR-NRGVKVKAIGKNKFDLIRRRNRKKSINFNYNNLEINNSIQYID 1658

Qy      398 ---NNILKNKSFDK-PREGFTSTFGKYSSLNDIDK-----IKKNKKK---GLINYKST-- 443

Db     1659 SMINENFKNKGFLKYKNDNLIISDANYVS-NNVDKDAASSYSMNNKKMDSKLLNNKDTNL 1717

Qy      444 ---------LYNDNTINKK------------------------HNNNNNNNNDNNDNN-N 469

Db     1718 VDKTDNVSVIYMDDSMNDKVSYSKSGIKENRGISNYAHITKNNGNNNNNDNNDNNDNNYN 1777

Qy      470 DNNNNNNNDSSSMVNNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYA---- 525

Db     1778 DNNNNNNNN-----NN------NNNNNNNNNNNNNNNNNNNNHDNNNNMDYYNYGGDKY 1826

Qy      526 FTDTSDFSSLDDMNCHLN---------------------------------NND---KSIF 550

Db     1827 NEKKDDFLNNDYQNYEKNYLKNDNKYDYDDNQNGFSTTLPRDMYEENQFYNNEEEKKRPY 1886

Qy      551 LHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNT----FF 606

Db     1887 RREKELGNNIHQGDKINQNNTYNNSYNSYNINNISTNIKMGKSSFNDTLNNKNSPSRLNY 1946

Qy      607 KSNINKMEDNTPKDILYESRNVSNMNGNVLLGL----------NKNTHHDIS-------- 648

Db     1947 KNNIEHDENN-----IMTSLNVNNLIESDLIDIKSNNFDYYKDNENNKYDGSMMLANYEQ 2001

Qy      649 -----TKD--ENHNDNKI----NNGVINIIN-NSNVNSIN------NSNMNSIN------ 684

Db     2002 KAQDMTMDVIDEYNNNYIEKESNKGNASIDNINKNINCINNDVDNINSNINNINDNIHKI 2061

Qy      685 NSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTS------NKNNDNI-FFDAI---SCDN 734

Db     2062 NSNVYGNDMSR-NYANNSNEKNVKINKKNESSLNTIDNLVNKNMQSINHFSSLMQSSKHN 2120

Qy      735 NMYP-------------------------NITNNNNNNNNNNNNNNNIDVEN-------- 762

Db     2121 KMSPNKGKLNVATDKNGSNEKNNMNGKNNMTDRNNMNDRNNMNSKNNMNSKNNMNSKNNM 2180

Qy      763 ---------YNNRDGTNNSMKLYAYNSHNLFQPD-------NNKNTSNIQNINTNK---- 802

Db     2181 TDRNNNNYYYNNRKGKNNEYDNYHNNNDEMTRKEISNDNFLNNENDSMKKEHNFNEFLEE 2240

Qy      803 --NNQDGNVNYS---------MNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRN 851

Db     2241 EYNNMNEHHHYSDVERIAGKKKNYSNNKTLDKLY--EYNNKYNEKRSNRNVDSN-VQMIN 2297

Qy      852 DFEKKKKTNFY-NNNNIVI---------VNNNMGNNNSPRMKYGLCGSHTSIDNMK---- 897

Db     2298 --SKRENVQFFEGNKNIDISFIPRTPHSLNKSIEENNGYNSDSNLLCSHNFVHNENEMLR 2355

Qy      898 --NNEMKNNEMKDNEMK----DNHIKSNNNNSSSSSSSN----NNIYNNI---------- 937

Db     2356 KINSQMSNYEDHDNKYKARNNSSNIYLNNNNNNNNSNSNYVHRNNVKPNIVKKFNRAKEE 2415

Qy      938 --NDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLN--------DYMSN 987

Db     2416 EDQEQEQEQDAIAHGKSGY---MKNNT----NVYKQEMIIENSNNMNKVNYKNNDDHKNN 2468
```

```
Qy        988 TLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDN 1047

Db       2469 MNITYLKKSHINQMSSFKE-KEKEKEKSNFRNPENNKY-SNKNNTSNTKFLSNIKSNSHN 2526

Qy       1048 NNVDNNVDNNDKNN-----------VDNNNVDNDDDDVDFHNIKNFNNNEYLS---YFQ 1092

Db       2527 NNNNMNAPNEQKNTNMVRIKGFMKGIQQNNKIPND--------IK-FNNDEYMNRKKNYE 2577

Qy       1093 KNVDTIINNCLNSLDISSMYDDTKEILN-NILLSKYKAEKD 1132

Db       2578 NKHEINSNNIIEQLD--------KTLMNKSIIMNKNKKKFD 2610

RESULT 27
Q8SSW3
ID   Q8SSW3     PRELIMINARY;     PRT;  1918 AA.
AC   Q8SSW3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Erythrocyte membrane protein PFEMP3.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115586; AAL92259.1; -.
SQ   SEQUENCE   1918 AA;  217043 MW;  BFBA078455C60D8E CRC64;

  Query Match            11.8%;  Score 807;  DB 5;  Length 1918;
  Best Local Similarity  21.1%;  Pred. No. 6.6e-25;
  Matches  361;  Conservative  204;  Mismatches  468;  Indels  674;  Gaps   55;

Qy        156 DLLQKRKMVAALENGKKEVVVKDLKILR-----VLTKEELILKMIDGVLLRKIGVNSQND 210

Db         29 EVIQTKSVQLRLQKEKEEGNIKMEKFKKRKRNQISDKED---------------INDLDD 73

Qy        211 ESSRSHAILNI-DLKDINKNTSLGKIAFIDLAGSERGADTV---SQNKQTQTD------- 259

Db         74 EDSNDDDDNNTNDDDDENQD-------------KQKGEETIKGFTSSKDVLGDFYKKVLS 120

Qy        260 ----GANINRSL-LALKECIR--AMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANIS 312

Db        121 GTPIGPSTSNSLNLGIPSIYQPFSTSSSANHKPFRSTSLSVSTFTPFKNDQQKPFLSPLK 180

Qy        313 PTISCCEQTLNTLRYSSRVKN---------FKNKSTCINEEDDTNTERIS--ILDSKGSE 361

Db        181 STLSTSSNTLPSMTSNSPEINKSIDPKDITHKRNSTG-NIPNNSNNSGLSNIIKNNISNE 239

Qy        362 MNASSIEN-------VVIKSNHLL-------SNNNNNKINRGKINDKIERNNIL--KNKS 405

Db        240 HNIISNNNNNTFFGPVTSKSNSLFGKNNNDNSNNNNNKNNNNDNNNDLANNNNSNNNNSN 299

Qy        406 FDKPREGFTSTFGKYSSLNDIDKIK---------KNKKKGLINYKSTLYNDNTINKKHNN 456

Db        300 INNSKNIINNNGNNLTSSNNNSNINISAGSNNSTKNTSTSNNNNNNNINNNNSNNINNNN 359

Qy        457 NNNNNNDNNDNNNDNNNNNN---------NDS---------------SSMV--------- 483

Db        360 NNNNNNNNNNNNNNNNNNNNFVGNNLSGGNNSLPNRVNPLSTSPIVNSSPITSPNQLSPR 419

Qy        484 -------NNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLD 536

Db        420 HNPIQAQNNNVNN--NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNIQNGNRNSFSGQI 477

Qy        537 DMN----CHLNNNDKSIFLHKKNLR---DNIKLKNRSSCDNIMNKKKN------------ 577

Db        478 SNNNNNGNNYNNNINSSNIGNNNVNGTTNNNNISNNISNNNINNKPSSQPIIQRVPNTQN 537

Qy        578 -------NLH-------------------------------------------------- 580

Db        538 IKQQPLPNIHQIQIQIPQSQVQQAPQQVPQQAPQQQQAQQQQAQQQLLQPQPQQNRISPQ 597

Qy        581 ------------------------------------LARHSVGSKLTMFSYDPQKNKDNTF 605

Db        598 PQQPQQQNRISPQPQQTQQAQQQQQPQQQAQQAQQQSQHQSQPQQNRISPQPQQPQTN-- 655

Qy        606 FKSNINKMEDNTPKDILYESRNVSNMNG-------------------------------- 633

Db        656 ---RISPQPQHQPTQINVQKRSIPQLDGINIDLDGKLIPILQSVPKEQIRQLQEQFVQMQ 712

Qy        634 ------------NVLL-------------------------------------------- 637

Db        713 QQQLQQQQQQQQLLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQRPN 772

Qy        638 -----------------GLN---KNTHH--------DISTKD------------------ 651

Db        773 QSQQYPQQNNIQSQPQQKGLTPPLKKVHTSNEQDDVVELITEDSKFSPISPNSSNNMGRN 832

Qy        652 --ENHNDNKINN---------GVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNS 700

Db        833 IEENVNGNFIVNGHNNRGRKSGPINI--NNNINN-NNINMNNNNNNNNNSNSSSSSNSNN 889

Qy        701 NQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNI---------TNNNNNNNNNN 751

Db        890 NIS-SNSNNNISNNNNNMNMNNNNVMNNSGRFSNNNGNNINIQGNGFNSNNNNNNSSNNN 948

Qy        752 NNNNNNIDVE-----------------------------------NYNNRDGTNNSMKLY 776

Db        949 NNNNNNISMNKNNMANTPLVIKSPNNSGNNQNNYINNGNNGNNNANYNNNNNNNNNNNNN 1008

Qy        777 AYNSHNLFQPDNN-----KNTSNIQNINTNK----NNQDGNVNYSMNFCHYNLN-DKN-- 824

Db       1009 NNNNNNNNNNSNNYNYHQNNNQNIQNNNNNKNINNNNQNISNNSSNNISNQNINTDKNRI 1068

Qy        825 ----------------YLIDLNNKEQKDKNIHGCD-----NNIIQNRNDFEKKKKTNFY- 862

Db       1069 NISVPNQPLSEHQQKILLIQAHQAQQIIKNEHTRQQQQFLQQQIENKNQSQQQVTKNHQQ 1128

Qy        863 --------------NNNNIVIVNNNMGNNNS----------------------------- 879

Db       1129 HLQLLIQQHQNQNPNPNQIQIQNPNSNQINQQLIQQYQQQQQHLLQQQHLLQQQQQQQQN 1188

Qy        880 -------------------------------------PRMKYGLCGSHTSIDNMKNNEMKN 903

Db       1189 QPPQQQQQPQQQNQPQQQKMQQLQPNQQPTVIQLPLPSQNYQRNSNQNSNQNPNQNPNQN 1248

Qy        904 -------------NEMKDNEMKDNHIK--------SNNNNSSSSSSNNNIYNNINDDDT 942

Db       1249 PNQNPNQNPNQYPNQYQNNPQSTHQLQKSQPTNNNSNNVNNANNMNNNNNMNNNNNMNNI 1308

Qy        943 FQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLS 1002

Db       1309 NNNNINHNNNNFNNNNNNNSNNNNNNNNNNKMNGGYNNYSNYQNNN-------------S 1355

Qy       1003 TNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNV-DNNNVDNNVDNNDKNN 1061

Db       1356 NNNNMNN-------------NNNNNNNNNNNNNNNNNNNNMNNGFNNNNNFNNNNFNNNKNN 1402

Qy       1062 V---DNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVD------------------TIIN 1100

Db       1403 IINNNNNNIRNSGDGPYYNSTNTVNNNNIRMEVKSNNEDNPVSSKPVMIGDTSFTLTLIN 1462

Qy       1101 NCLNS-------------LDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNM 1146

Db       1463 DFVRPFPEQVIALKETTFYLRINVSGGAANSVINQDTLKRLSGK---AVVKGTHEDITAT 1519

Qy       1147 SLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMC-------HICN 1199
```

```
Db        1520 SPNDSFKLHQDVKSEQGVL----------QFDVKIRNQGVNTPSHLVLIFHLIDHGSIVH 1569

Qy        1200 NNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSS 1259

Db        1570 SIESKRISFFTRSKYAPYLPIFYIMQSKNCFLPLSRSLYR------NNTESNTNTNNSSS 1623

Qy        1260 NNGDIILLNKKLVQDNIKNSMDHNNIH 1286

Db        1624 SN-------------NNSNNNNNNTY 1637


RESULT 28
Q8IBU8
ID   Q8IBU8      PRELIMINARY;      PRT;  2206 AA.
AC   Q8IBU8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF07_0061.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA   Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844506; CAD50900.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2206 AA;  258401 MW;  D4ED711AE470383C CRC64;

  Query Match             11.8%;  Score 804;  DB 5;  Length 2206;
  Best Local Similarity   22.7%;  Pred. No. 9.9e-25;
  Matches  353;  Conservative  226;  Mismatches  381;  Indels  598;  Gaps   73;

Qy         168 ENGKKEVVVKDLK--ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKD 225

Db         467 ENVKEDIIKEDIKEDIIKENVKEDIIEEDIKEDIIKE---NVKED-------IIEEDIKE 516

Qy         226 INKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLA--LKECIRAMDSD-KN 282

Db         517 -------------DIIKEDVKEDIIKENVKEDIIEEDIKEHIIEEDIKEHI--IEEDIKE 561

Qy         283 HIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCIN 342

Db         562 HIIEED------IKEHIIDK-------------CVEQDI----INRNISNEINNSYLLN 597

Qy         343 EEDD--------TNTER-ISILDSKGSEMNASSIENV---------VIKSNHLLSNNNNN 384

Db         598 ENIDKIKTSKEGSHEERPRTVLKQIPANISCCPINSIEQEKYNNDDSLKSTLSCSVNNNN 657

Qy         385 --KINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNK----KKGLI 438

Db         658 MKKKTTLKLSEPSEEVNMNKQK----------------KENETEKIKGNKIECERK--- 697

Qy         439 NYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNN 498

Db         698 --------NRRVSEQIDDNDNNNNNNNNNNNNNNSNNNN--------------NNNNNNN 735

Qy         499 INVNNNNNNNNNNNNSHNN-HLPQPNYA-------------------------------- 525

Db         736 NNNNNNNNNSNNNFNSNNNVVLPNNNYLLNNKLCIPIDDEYIKKIRNEQTDTFLQKLIME 795

Qy         526 ----------FTDT------------------------SDFSSLD-------------- 536

Db         796 YYDERKRRTFFDDTIKGNIYYVNDYEEYREEFESNRTKIGDLQNFDKFILPNQFILSSNE 855

Qy         537 ------------------------DMNCHLNNNDKSIFL-----HKKNL----------- 556

Db         856 TDINENKYNDIKSSNTKNNDQQNKNVNKHNNNNNKLIHITTLENNNKNVTLKEGQHIHIS 915

Qy         557 RDNIKLKNRSS------CD-----NIMNKKKNNL--HLARHSVGSKLTMFSYDPQKNKDN 603

Db         916 KDSMHISNKKNIKHTVECDEDNFLSYKNKIKNDILQELVNKSFNSNSINDIYNNKENKIE 975

Qy         604 TFFKSNINKMEDNTPKDI--LYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINN 661

Db         976 -LNDSHLKHCEVNISNNMREYNQSRNMEHV--NVLVHMNNN-------NDKDKNNNERNN 1025

Qy         662 GVINI--INNSNVNSINN--SNMNSINNSNMNSNSIYK----SNYNSNQSISDVQIRYVN 713

Db        1026 NVCNLKNISSFNVNTMNNMINNNNNNNNNNINFNSPNKYYNFNNINGEFSLND------- 1078

Qy         714 EMDTSNKNNDNIFFDA-----ISCDNNMYPNITNNNNNNNNNNNNNNN----------- 757

Db        1079 ----DRRNKGNIPFDMNLTNDISKNNNNNNNNNNNNNNNNNNNNNNNNNNKIMDTSLKS 1134

Qy         758 ------------------------------------------IDVENYNNRDGTNNSM 773

Db        1135 MRAVSLSSLQKLILEHNKILREKHNHKITQIGNQDNCIFKISESAESENENNEHNNNNNL 1194

Qy         774 KLY--AYNSHNLFQPDNNKNT--SNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDL 829

Db        1195 NIIHSINNTNNIHFPNNKVNTFVSNKTNVNMN-----------NITNINMNNVINSI-L 1241

Qy         830 NNKEQKDKN------IHGCDNNIIQN---------RNDFEKKKKTNFYNNNNIVIVNN-- 872

Db        1242 NNAANNNMNNAPTNVMHNTSNSLMNNAPTNVINNTSNSLMNNAPTNVINNTSNSLMNNAP 1301

Qy         873 -NMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKD-------------NEMKDNHIK 917

Db        1302 TNVINNTSNNL---MNNAPTNVINNTSNSLMNNAVSQLINNGTNHMINNGTNHMMNNILN 1358

Qy         918 S--NNNNSSSSS-----SSNNNIYNNI-NDDDTFQ-------------NDYCHNDNTFT 955

Db        1359 SLINNNNMMMMNIPVDIDKNSKVFKNIENEEPVLQNLSDIQKLLLHKGSNPYMFNNLRIS 1418

Qy         956 --IRRKNNTNINSNIYQN---DDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNK 1010

Db        1419 NNLNNDKNSNVNSFIKMNNVGENISTCQHNNNNYMVN------------ITTSQPRYNC 1465

Qy        1011 EMEGKHIRLDDQDKYDDNDNNN----VDNNNKNNVD-------NNVDN-NNVDNNVDNND 1058

Db        1466 -INRYNVPKDYEENKDINKKSNVIKKVDNEKCYTIDEKYGGPYNNCENVNDMKHRRGNSN 1524

Qy        1059 KNNVDNNNVDNDDDDVDFHNIKNF--NNNEYLSYFQKNVDTIINNCLNSLDISS------ 1110

Db        1525 CSNT-NNNKDNKDNYGNYGNYDNYDNNNNNYMNYHNNNINDYANHHFNKQHVNTIREGFD 1583

Qy        1111 ---MYDDTKE-------------------------------ILNNILLSKYKAEKDNVI 1135

Db        1584 TYILIDDDNDNDTDISTDKRNNPLNRNNKKMNLKCNLNKLPNITNN---EKHKNEKNISR 1640

Qy        1136 KKYI--------NEDIKNMSLEEIDKTAQSI--YEKRK---------VLLTKLLLLFKKN 1176

Db        1641 TKYVCLGLKRNSRENIHHMNNEENEINLEMINDYHQNKQVKDIISTYIHTSKIQYETKKG 1700

Qy        1177 VDTQINN----------------ETSDLR-----KDLVMCHICNNN------------- 1201

Db        1701 KDISNNNYMNVQKIQTTHGNYPHKIGDINIQENYRNVVQNRNMNNNSSFAENIGRILTRS 1760

Qy        1202 PDDQFHFYAYSRLEKDIINLIMLRQIW---------------CESENLRLLYQFLVVEYQ 1246

Db        1761 SSQQKDIHA---LEKEIGNVQNDKSIYYNNTVNRWNDEVVFICTNEN----------EDK 1807

Qy        1247 NKSANSVLLNVSSNNGDIILLNKKLVQDN----------------IKNSMDHNNIHKK 1288

Db        1808 NDSANS--LRKEKKNGTLLYAKEKEEENNTETQNVRRSVRLHNKKVKNETKNREIEKK 1863


RESULT 29
Q8IB94
ID   Q8IB94      PRELIMINARY;      PRT;  8591 AA.
AC   Q8IB94;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Ubiquitin-protein ligase 1, putative.
GN   MAL8P1.23.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA   Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844507; CAD51113.1; -.
KW   Ligase.
SQ   SEQUENCE   8591 AA;  1005145 MW;  76B6C5F8AFEB8B1C CRC64;

  Query Match             11.8%;  Score 804;  DB 5;  Length 8591;
  Best Local Similarity   25.5%;  Pred. No. 3.5e-24;
  Matches  320;  Conservative  205;  Mismatches  419;  Indels  312;  Gaps   58;

Qy        206 NSQNDESSRSH-----AILNIDLKDINKN---------TSLGKI---------------- 235
              |: |: ::||:      |:|||  ::||  |         :| |
Db       5393 NNNNNNNNRSNNNSFSNIVNIDAQNINNNQMRNRMNTTNDIGHINNLQENINTNHMNSTS 5452

Qy        236 -------------AFIDLAGSERGADTVSQN--------KQTQTDGANINRSLLALKEC 273
                           |      |     |  :  |          :  |   :|| |||
Db       5453 QYNMNGNISNICNNTFAGQKSSSNERDVLENNVNNSIIIHNENSTHDHHINSSLL----- 5507

Qy        274 IRAMDSDK-----NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQT------- 321
                  |::|      |:|     :    :   :|     :: :  ||     :|
Db       5508 ----DNNKNQNISNNISMIHNSTIETKNNILPNALENNIHNNIDTQNNLCNNVNANSCSN 5563

Qy        322 ---LNTLRYSSRVKNFKNKSTCINEEDDTNTERIS----------ILDSKGSEMNASSIE 368
                  ||     | :      ||  :. |  |::|| |  |:              ||::|    :     |
Db       5564 NIYANTQTKSEKKGTIKNGISLENNEENTN-EHINKNNVIVGENYILNNK---IKNEQIN 5619

Qy        369 NVVIKSNHLLSNNNNNKIN-RGKIND---KIERNNILKNKSF---DKPR-EGFTST---- 416
              |  :|::|:: :|  ||  |   :  |||    |::|||  | |    ||   |    ||
Db       5620 NTKVKNDHIIDSNINNINNIKNDINDNIIKVDKNNIKNNDSFAVKDKSMLESSESTTNIR 5679

Qy        417 -------------FGKYSSLN-DIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNN 462
                           :  | | |  ||: :        |    ::    |     |  : :|  |||||
Db       5680 NVDENNVSIIDLYYHFYGSNNRDINTLLNENNSILYLNETGQINSTQIKETNNKINNNNN 5739

Qy        463 DNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNN------INVNNNNNNN----NNNN 512
              :||:|||:||||||| ::: |:|  :   : : | |      : :||||         ||
Db       5740 NNNNNNNNNNNNNNNNGNNTSVDNKDSKKTDKSNNTNDAEQVALQLNNNNTEGVEEFNNIT 5799

Qy        513 NSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIM 572
              : :  :|    |:   ::|   :  |         | ||:    :  :|: : ::   |     :||:
Db       5800 SLNIENLMDQLYSTLLSTDIRTND-----LLNNE---IISLQNIANRVQRNN--DFNNIV 5849

Qy        573 NKKKNNLHLARHSVGSKLT--MFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSN 630
              ::      |:        :|  :|        || : |::|:                    : |   |||
Db       5850 DESSYNI------IGDDITENRNLYDNRLNRNNS---------------VPYIHRNVER 5887

Qy        631 MNG-NVLLGLNKNTHHDISTKDENHN-DNKINNGVINIINNSNVNSINNSNM----NSIN 684
              :|   |   | | ||:   |  :  | |  | | |  | |:|   |:  :|| |:|:    |:|
Db       5888 LNNFPPLPTSNNNTNIRASWNNRNINRPNSITNNVVNNSLNNITSSIRNNNLRNIFNNIR 5947

Qy        685 NSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNN 744
              |:| | |        :|  :::|: |               |   | | |    :  |  ||  || ::::
Db       5948 NNNNNENV-----FNRHRTINIV---------PSIIRNDIIQNLSASHQNNRDPNSSSSS 5993

Qy        745 NNNNNNNNNN-NNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKN 803
              :||||||||   |   |   |  |   ||                           || |  | |:|
Db       5994 SNNNNNNNNNLTTNRDTNNTNERTNENN---------------------NIHNNNNNEN 6031

Qy        804 NQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNF-- 861
              : : |: : :          | ||:  :||  ||||               | ::   :||
Db       6032 DNNNNISQT----RVNNNDQRQVSDLN----KDKN---------QIKDKGSKIEKTEMLT 6074

Qy        862 --YNNN----NIV--IVNNN--MGNNNSPRMKYGLCGS----------HTSIDNMKNNEM 901
              ||  |      |:|  :::|:  : :       ||    || :                |:  |
Db       6075 KKYNKNQAQFNMVDAVIDNDTILKSAICNRMYNRLCKAIPKVHIKLPLENLIEEKKRIIQ 6134

Qy        902 KNNEMKDNEMKDNHIK--SNNNNSSSSSSNNNIYNNINDDDTFQNDYCHNDN------- 952
                  : |: |      :||   ||||:::|:  :  |     ::| :      ||        |
Db       6135 TLKKQKEQEQLQENIKGNDNNNNNNNSNKEHQNELEDVNKNKGDQNKKDKKGNKDKKLSR 6194

Qy        953 --TFTIRRKNNTNINSNIY---QNDDIIYT-INSLNDYMSNTLL----HFKEKYTYPTLS 1002
                | |    :| |||   ::|        :||:|        : |  |: :  || |       :       |       :
Db       6195 FTTTTNDKKKNTNEENDIIESNENDNITNAQLGSHNEQLENTELNNDNNNNNNTNNGFN 6254

Qy       1003 TNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNV 1062
                ||     : : :                | |  |  |||:  ||||          |||| |||  :||| ||
Db       6255 PNEGDNDSDSDSNSDSNSDSDSDDYNDNS--DNNNNN------DNNN-DNNDNNNDNNNE 6305

Qy       1063 DNNNVDNDDDDVDFHNIKNFNNNEYLSYFQK--NVDTIINNCLNSLDIS-SMYDDTKEIL 1119
                : |||:||:|  ::    | :::  |      |         |::::: |    || || |  :  :     :
Db       6306 NQNNVNNDEDHLEEENEEDYKNINYEELFDNILNIESLLIICC-SLGISESQFLELLNVD 6364

Qy       1120 NNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKL------LLLF 1173
                : |:          |:|  :  | : :||::|  | :  ||    : :|::   |:         :
Db       6365 RSFLIEL----PSNLIDELIIQHLKNITLTGISEKAQDNIKNKKIISIKMYKNYLKYQEY 6420

Qy       1174 KKNVDTQINNE--TSDLRKDLVMCHICN-NNPDDQFHFYAYSRLEKDIINLIMLRQIW-- 1228
              ||: :   :||   |:|: :          :|  || |:  :|        :   : :| |  :
Db       6421 KKSQENNASNENNTNDILNNTTGNNITNDNNNNDNNN----NNNDNNINNNVSSANSTTN 6476

Qy       1229 -CESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKLVQDNIKNSMDHN 1283
                | :|            |  |   :||: |: :|                 :|:|| :  : |
Db       6477 NCATE---------FNERPNNENSSVIQNLDNNTN---------IQNNINDEKEQN 6514

RESULT 30
Q8I5Z0
ID   Q8I5Z0     PRELIMINARY;      PRT;  1785 AA.
AC   Q8I5Z0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFL0275W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014844; AAN36144.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1785 AA;  204851 MW;  E07C24F61F078282 CRC64;

  Query Match             11.7%;  Score 803;  DB 5;  Length 1785;
  Best Local Similarity   25.0%;  Pred. No. 9e-25;
  Matches  309;  Conservative  202;  Mismatches  385;  Indels  338;  Gaps   55;

Qy        135 DKDNTKGIFISFYEI--YCGKLYDLLQKRKMVAALENGKKEVVVK-DLKILRVLTKEELI 191
                : : :  |||:  :|  ||                        ||  :    :::| |       ||||:
Db        783 ETNSNQDIFITEDQIPSYC---------------ENIESNLLIKFDFG----TTKEEIT 822

Qy        192 LK--MID--GVLL----------RKIGVNSQNDESSRSHAILNID-LKDINKNTSLGKIA 236
```

```
Db      823 PKTEYIDPKGVVLGRGPYAQSSYKKISVTNSNGYVSREHCLIYYDGSKDAGERWLL---- 878

Qy      237 FIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLR 296

Db      879 ----------------RDTSTLG-----TFLKIKPF-------SNPIPLPIGSIFK--- 906

Qy      297 DIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILD 356

Db      907 ---AGQCKIEV---------CSHDSAQFIQYPERSISLHNTNNPPNTDNN---------- 944

Qy      357 SKGSEMNASSIENVVIKSNHLLSNNNNKINRGKINDKIERNNILKNKSFDKPREGFTST 416

Db      945 ----------------------NNNNNADNRNNDNDNNNTNNNNNNNNNNNNGGSNNS 982

Qy      417 FGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNN 476

Db      983 SG---STNN------------NGGNNSANNHTNNNNNNNNNNNNSNNNSNNNNSNNNN 1026

Qy      477 NDSSSMVNNMINH-MINNNINNNINVNNNNNNNNN-NNNSHNNHLPQ--PNYAF----TD 528

Db     1027 NNNSNNNSNNNNHSSSNNAPNSSHANNNHSNDNNINNNYRNNHVHTCISNNSYHRNSGN 1086

Qy      529 TSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGS 588

Db     1087 TNNFS--NNMHSSYNNIINGIY----NYQNNSRNLNNSSSFGRTNILQDNSH---HENMS 1137

Qy      589 KLTMFSYDPQKNKDNTF-------FKSNINKM-EDNTPKDILYESRNVSNMNGNVLLGLN 640

Db     1138 NMA--------NQNNNFEAIVQNLFQRNVNTLGRENNPS--LFSSRNEQ---------LR 1178

Qy      641 KNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNY-- 698

Db     1179 LNNNHNIMNNNNTNNNNNNNNN-----NNNNNNNNNNNNNGNNNNHCSHGDTYFYNYLR 1233

Qy      699 --------NSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNN--- 747

Db     1234 YSVNNNGRNHNNSGS---IDYLNDNEESDNNNNN---------NNNNNSSTKERNNNRSS 1281

Qy      748 ---------------NNNNNNNNNNIDVENYNNRD----GTNNSMKLYAYNSHNLFQPDN 788

Db     1282 VNDSCIVENHQHYHQNNNNNNNDNNNDEHDHNNNNSSGHGNNNNNN---SNSHSNSQNNN 1338

Qy      789 NKNTSNIQNINTNKNNQDGNVNYSM------NFCHYNLNDKNYLIDLNNKEQKDKNIHGC 842

Db     1339 NNNNS---NSNSNHNHGNGHNNYNKSHSNNHNYNHLNVNTPDYPFNVNTRFREHAYNSSL 1395

Qy      843 DNNIIQNRNDFEKKKKTNFYNNNNIVIVNN--NMGNNNSPRMKYGLCGSHTSIDNMKNNE 900

Db     1396 GSCFLKNKNSEMQIPQLNCTLNSNHNNMNNFVNCNDNNYNFIPYNL---------------D 1442

Qy      901 MKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNT------- 953

Db     1443 QSNSKYENNISFINNNNNNNNNYNNNYNNNNNMINMINMNSLNNQHTAALDNTNSLFNID 1502

Qy      954 -FTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEM 1012

Db     1503 KFTLERR-NINICANI---NSIFNNINRDSPCLDNRV----------NLASNETIQNAPI 1548

Qy     1013 EGKHIRLDDQDKYDDNDNNNVDNN-NKNNVDNNVDNNNVDNNVDNNDKNNV--------- 1062

Db     1549 RNMH-SLNVNKKSEKPKTNDVYKNINKNNIERKNSKNGLKKFLINNNKLDILLKKKRNHS 1607

Qy     1063 ------DNNNVDNDDDDVDFHNIKNFNNNEYLS-YFQKNVDTIINNCLNSLDISSMYDDT 1115

Db     1608 ETFLYTNQMSTTQDEEETGPNTINTTNYAEAGGIYNEMNSLKLIYNYINNLNINNKKTNV 1667

Qy     1116 KEILNNILLSKYKAEKDNVIKKYINEDIKNMS----LEEIDKTAQSIYEKRKVLLTKLLL 1171

Db     1668 SP-TNNII------ECNSDLKLENNANTKNLNSLICSGEFNHITENNHNNNN-------- 1712

Qy     1172 LFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHP----YAYSRLEKDIINLIMLRQI 1227

Db     1713 --NNNMNSMINNINSH-------NHIYNNNQNNINHFNDLHHSYNNINDPI--------I 1755

Qy     1228 WCESENLRLLYQFLVVEYQNKSANSVLLNVSSNN 1261

Db     1756 IGDIKTKKLLFS----DHNNVNQNSFFQSIKMNN 1785
```

RESULT 31

```
Q8II48
ID   Q8II48      PRELIMINARY;      PRT;  2763 AA.
AC   Q8II48;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF11_0326.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014840; AAN35910.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2763 AA;  330300 MW;  23AF0D9286193774 CRC64;

  Query Match             11.7%;  Score 802;  DB 5;  Length 2763;
  Best Local Similarity   24.8%;  Pred. No. 1.5e-24;
  Matches  316;  Conservative  166;  Mismatches  358;  Indels  434;  Gaps   59;

Qy      248 DTVSQNKQTQTDG-ANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSI 306

Db       11 NTSNNNKKNNTNNKINNNTNNTSYSNGVKNLDS-------TNPEQNKNLR---------- 53

Qy      307 MIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASS 366

Db       54 ---SVSNNMSCSRHTDICNNY-----KFHNNN---NKEEQTNIFGKNVTSSNDQYNKTDR 102

Qy      367 IENVVIKSNHLLSNNNN---NKINRGKINDKIERNNILKN--KSFDKPREGFTST----- 416

Db      103 NNNGRDKKNHARNRKNNRRDKKNNHGKVDGESNNNNNYNNMIKSMCENNFGIQQNNVHNM 162

Qy      417 --FGKYSSLNDIDKIK------------------KNKKKGLI-NYKSTLYNDNTINKKH 454

Db      163 DWFFKNENLCNMNNVRDQNGESFLCNNKGINMLQDKSYNKGIFKNDKNKIMNPDNNNNNN 222

Qy      455 NNNNNNNNDNNDNNNDNNN------------------------NNNNDSSSMVN----- 484

Db      223 NNNNNNNNNNNNNNNNNNNIYRNPYVEENNIMAHKYNLSLDQRKNNKNKNTTYINKNKVF 282

Qy      485 ----------------------------------------NMI----------------N 488

Db      283 EDHEMKNSLQKKKIYDETIPVGNFVQRKKNYKKDGDDIKNMMVSGPKFSYTEKVYKNDNN 342

Qy      489 HMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKS 548

Db      343 H--NNNNINNINNNNNNNNNNNNNNNNNNNNCTYNAHKYSDIYNIGDEQNHIINNDHP 400

Qy      549 IFLH------KKNLRDNIK--LKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600
```

```
Db        401 YFLNKYKGKGKDTKKDSVRSTLPYNNICTNDEKKEMN--------ITKKHQTHEYTQDTN 452

Qy        601 KDNTFFKSNINKMEDNTPKDILYESRNV-SNMNGNVLLGLNKNTHHDISTKDENHNDNKI 659
              |                      | ||  ||  |     | :|::    :: |  |
Db        453 K----------------------SYNVMDNMFINYFKNKNASTNNVFELEENNRGD--- 486

Qy        660 NNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSN 719
                : ||:|| :::  |:  | |: ||:|  :::  : :::|||    :|          | ||
Db        487 ---ICNIVNNIDIDKNNDHNNNNDNNNNDDNSKDHSNDHSNDHNND---------DDSN 534

Qy        720 KNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNN--NNNIDVENYNNRDGTNNSMKLYA 777
              :: |          ||| |   :|| |||  || ||   || ::    |::   |  |:|
Db        535 DDDSN------DDDNNNYYYYSNNYNNNYRNNKNNFKRNNFEI----NKE-KENKEKIYT 583

Qy        778 ----YNSHNLFQPDNNKNTSNIQNINTNKNNQDGNV-NYSMNFCHYNLN---DKN----- 824
                  |:  :        ||||  |   |||  |:|:  |: ||:  :    :::   |||
Db        584 SVFEENNKVIIDMINNKNKDNFININ-EKSNEITNIHNYTDVYVDEDIDIHVDKNNNIFF 642

Qy        825 ------YLIDLNN-------KEQKDKNI----HGCDNNIIQNRN----------DFEKK 856
                    :: | ||        |:  |.||       :||     |            ::| :
Db        643 DHNSNFFVDDFNNLHVYNKHNEEYYKNIKLYNRFMENNKTGKENYDYIQEYEYEYEYEDE 702

Qy        857 KKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCG----SHTSIDNMKNNEMKNNEMKDNEMK 912
                :| || :    | |: :   ||   :: :  :         |  || || ||:  |:   ||
Db        703 DLSNNYNMD--YIYNDQIQRNNDSQIIHDINNYDIFSILSIPNMNNNDNNNDNNNDN--- 757

Qy        913 DNHIKSNNNNS----SSSSSSNNNIYNNINDDDTFQNDY---CHNDNTFTIRRKNNTNIN 965
               |:  :||||     ||||| ||| ||          :|: |       |:||:  |    ||
Db        758 -NNDNNNNNNDNYYYSSSSSKNNNYYN-------YQDRYIPHLHHDNSHT---NNNINNI 806

Qy        966 SNIYQN-DDIIYTINSLNDYM---------SNTLLHFKEKYTYPTLSTN--------- 1004
              :|| :| : ::    |   | ::            |    |  :||    :     |
Db        807 NNIIENYNPLLIYHNEWNHFLKDQKEKMKNTDNIKKHENDKYNIKSHDNNFLQEIMNKNF 866

Qy       1005 --EDI--YNKEMEGKHIR----------------------------------------- 1018
                |||    ||:    |:|
Db        867 IYEDIKTTNKKYFEKYISDNVVPKTHPLDYIINNDMNYVQKEDTNNNPFHDKTDKGTIQK 926

Qy       1019 ----LDDQDKYDDNDNNNVDNN-----NKNNVDNNVDNNNVDNNVD------NNDKNNVD 1063
                  ::: |  : :|  ||: :|       :| |||||:    ::: | |          :| ||:
Db        927 IYDLINNHDDHINNKGNNISHNFISDHSKMNVDNNIHVEDINQNGDDIKCSKSNKTNNIF 986

Qy       1064 NNNVDNDDDDVDFHN--IKNFNNNEYLS-YFQKNVDTI--------------INNCLNSL 1106
              |||   :| ||  |      :| :|   :|   | |:                    |||  | |
Db        987 NNNKKDDYDD-DNHKCIFQNKDNTNHLDIYKQSNISNTHIEEHGKEHLNKGDINNTKN-L 1044

Qy       1107 DISSMYD-----------------DTK-EILNNILLSK-YKAEKDNVIKKYINEDIKNM 1146
                  |: ||                    |    :  ||  |:  : :||:      :|  |
Db       1045 HASNKYDALHKECTIYIDNISYYGQDNKIDYTNNKKDSQGMEKKKDHTTNIQLNHLLVN- 1103

Qy       1147 SLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN------ 1200
                |:  ||:  || |  | : | ::           ||              ||:  ||
Db       1104 --EKNKKTSSIIYNKGMVNINKNMICNNNN-----NN---------MCYPINNCSNEYD 1146

Qy       1201 ---NPDDQFHFYAY 1211
                 | | ||::| :
Db       1147 KKDNKDHQFNYYKW 1160
```

```
RESULT 32
Q8I3A3
ID   Q8I3A3      PRELIMINARY;      PRT;  1770 AA.
AC   Q8I3A3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFI0225W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22255708; PubMed=12368867;
RA   Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA   Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA   Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA   Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA   Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA   Feltwell T., Goble.A.; Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA   Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA   Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA   Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA   Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA   Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch E.,
RA   Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA   Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA   Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA   Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT   "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL   Nature 419:527-531(2002).
DR   EMBL; AL929355; CAD51731.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1770 AA;  207567 MW;  83195F911DFC548F CRC64;

  Query Match             11.7%;  Score 801;  DB 5;  Length 1770;
  Best Local Similarity   23.0%;  Pred. No. 1.1e-24;
  Matches  363;  Conservative  236;  Mismatches  476;  Indels  506;  Gaps   78;

Qy         12 RPLSELEKKKKDSDIITVKNN--------------------CTLYIDEPRYKVDMTKYI 50
              |||::            ||      |||                     |  :   |: |  :  || :||
Db        239 RPLNDNSNNISKDDIKEKKNNNINSTVNYDNTNTEENITSDHCNIK-DDTRLEKDMEEYI 297

Qy         51 ERHEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQP 110
              ::                        :| :      :| :|| |                    |
Db        298 KKK--------------NIYMSNSNKIINELYNN--------------------LIYDE 322

Qy        111 YGQS--DTPGIFQYAAGDIFTFLNIYDKDNTKGIFISF-YEIYCGKLYDLLQKRKMVAAL 167
              | ::       |: :     : :     ||: | |  :    ||        |  :|| :
Db        323 YSENILSKKGVKEKDHIEYYEEQNIHMKANEESTNISIDIPPCCQIIYDNVDDATN-EQY 381

Qy        168 ENGKKEVV------VKDLKILRVLTKEELILKMIDGVLLRK-IGVNSQNDESSRSHAI-- 218
              :| :|:                :  |:|   :  |   :  |||  |      :|  |  :    ||  :    :  |
Db        382 DNSQKDTYNWYMQKTNNNKLLYHINKNLIFLKRIQQYFYQKYINIKFSNDTNDYYYIIHL 441

Qy        219 --LNIDLKDINKNTSLGKIAFIDLAGS-------ERGAD------TVSQNKQTQTDGANI 263
                  |  ||   ::            |  ||         |      |         :|::     :|   |:
Db        442 EWFNKLKKFINNESN-------DFPGSISNWELYEYTHDEIFKNYNISESNYVFSDDKNM 494

Qy        264 NRSLLALKECIRA-MDSDKNH-------------------------------IPFRDSE 290
                |  ::       ||:|::       |  |::                                 ||        ::
Db        495 NDNIYLKKQCLKKNLKEGKDYICTNKYMWRFLQFLYNGGPCIKRISNNIYNTFIPISSND 554

Qy        291 L----------TKVLRDIF-----VGKSKSI-----------------------------MI 308
              :          :: ::::|        :  :| |                                   ::
Db        555 IMNNNIMYLLESRYIKNLFSLFNYIDHTKFIYNEPKGNEHTLYKNEYYNDNDKYTHDYIL 614

Qy        309 ANISPTISCC-------------EQTLNTLRY------------SSRVKNFKNKST---- 339
                :     |                  ||    |  :  |                  :|   |||::
Db        615 EETNEKKMCAHNYHELLQFYNLKEQEKNIILYIEYDDKHINKEILDEIKKIKNKNSNNKQ 674

Qy        340 ---CINEE------------------DDTNTERISILDSK--------GSEMNAS---SI 367
                    |:|                         |    ||::: :|::                  | ||  :      |:
Db        675 NILISNDENFSSDSSNMYNIINAKHNDKLNTQKLFLLENDKICANSHISSNMNQTEYISL 734

Qy        368 ENVVIKSNHLLSN-------------------NNNNKINRGKI--------------NDK 394
              :|         :::||:|                    || |   |  |  :                     |:
Db        735 DN--FDADYLLNNPHNLSRGFPNSYKLDINTDNNENVDNNGNVDSNENVDSNENVDSNEN 792

Qy        395 IERN-NILKNKSFDKPREGFTSTFGKYSSLNDIDKIK------KNKKKGLINYKSTLYND 447
              :: |  |:    |::  |        |         |:: |  :|   |          |:::   :       ||
Db        793 VDNNENVDSNENVDN-NENMDRNDNMYNNEN-VDNSKMFINCNKSQRSNIKKSNSTNSTR 850

Qy        448 NTINKKHNNNNNNNNDNNDNNNDNNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNN 507
                   |: :||||| ||:||:|||:|||||||::|        ||           ||| |||  | ||||·|
Db        851 RNYNRNNNNNNNKNNNNNNNNNNNNNNNNNNSRNNNN------NNNNNNNNNNNNNNN 904
```

```
Qy        508 NNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSS 567
Db        905 NNNNNNNHNNNNKDDN-----TSD-----------NNNN--------NINKEEDKKNKTT 940

Qy        568 CDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRN 627
Db        941 ----NNKKKEN-------------------EKDDENK-CKGNLNGSVE------IYELKR 970

Qy        628 VSNMNGNVLLG---LNKNTHHDISTKDENHND-NKINNGVINIINNSNVNSINNSNMNSI 683
Db        971 EFEENNNIIYNDSYNNRNINNVIDDLKKNEEDINKINDRNIYLSPNISANEMNINNFNKS 1030

Qy        684 NNS-----------NMNSNSIYKS--NYNS-NQSISDVQIRYVNEMDTSNKNNDNIFFDA 729
Db       1031 YNSSSNKKNSIPCNSSNGNDIYKSCEEYNNDNEKIS--SNGYLTTTESQSKGTTD----- 1083

Qy        730 ISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNN 789
Db       1084 ---GNTERGSIYEYENDNNNNNKNNNKN---ESNNNR----NEKKIYYDSIENL--DDVV 1131

Qy        790 KNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDK-NYLID---LNNKEQKDKNIHGCDNN 845
Db       1132 KKRKHIKNA---QNNTTNNRVCSSNCGEQQVTEKINNILDNTHLNNIQNKNHNLKNNNSS 1188

Qy        846 IIQNRNDFEKKKKTNFYNNNNIVIVNN--NMGNNNSPRMKYGLCGSHTSIDNMKNNEMKN 903
Db       1189 TIQNGCTI-KGNEQNVKNTNNINEEDNITNLENHKK-------------DQKKKNHIMK 1233

Qy        904 NEMKDNEMKDNHIKSNNNNNSSSSSSNNNIYNNINDDDTFQNDY-----CHNDNTFTIRR 958
Db       1234 KKLDDIDVKQGDLKSNNH------ENKNDVEDNMEMDETNNNSMDPQQRCNLISVFNKQK 1287

Qy        959 KNNTNINSNIYQNDD-----IIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEME 1013
Db       1288 NHKNNISNNNNKKDDDDDDQSVYSSNITNTNSSS--LH--NSCSSSSSGGNNSLYNE--- 1340

Qy       1014 GKHIRLDDQDKY---DDNDNNN-----VDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNN 1065
Db       1341 ------NDISKYNIFNNNDNDNLKNLLVPNNNSNN--NNNNNNIIIINSNNNNNNNNNNN 1392

Qy       1066 NVDNDDD-DVDFH----------NIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDD 1114
Db       1393 NFKRDNESSLNYHTSIMTKEQPAGIINYSTTCYIN--------VVMQCL------SVFFK 1438

Qy       1115 TKEILNNILLSKYK------AEKDNVIKKYINEDI------------KNMSLEEIDK--- 1153
Db       1439 LIYTLHNYVTVKYKNVNMSSDENENMNSSFINKNFFTNSIPFNIFGSNNNNNKKKDECLL 1498

Qy       1154 --------TAQSIYEKRKVL-LTKLLLLFKKNVDTQIN-NETSDLRKDLV--------MC 1195
Db       1499 LTFSFKLFQLSKMHNKGKVLCVNKLLNLLNDKYSYLFEYNEQQDCHEFLLLVFDFIHNMV 1558

Qy       1196 HICNNNPDDQFHFYAYSRLEKDIINLIML----RQIWC-ESENLRLLYQFLVVEYQNKSA 1250
Db       1559 KVIDESVDKNNQIDYYLKKEQSIISDLFLGLIEEKITCSQCEYVNYIYQ----PVYNLSV 1614

Qy       1251 NSVLLNVSSN-NGDIILLNKK 1270
Db       1615 NVFKKNPENNINDNLIEYFKK 1635
```

RESULT 33
Q8IHP9

```
ID   Q8IHP9      PRELIMINARY;      PRT;  2940 AA.
AC   Q8IHP9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF11_0480.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014843; AAN36060.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   2940 AA;  348040 MW;  BF8748D8A051BAD2 CRC64;

  Query Match            11.7%;  Score 800;  DB 5;  Length 2940;
  Best Local Similarity  23.8%;  Pred. No. 1.9e-24;
  Matches  360;  Conservative  233;  Mismatches  444;  Indels  478;  Gaps   77;

Qy        132 NIYDKDNTKGIFISFYEIYCGKLYDLLQK---------RKMVA--ALENGKKEVVVKDL 179
Db       1116 NSNSNSNQVCLYMPNHELY-NSNYNITQHMPSSNDVPKIQKHIANNIIMNGHKEEHIIEK 1174

Qy        180 KILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINK-NTSLGKIAFI 238
Db       1175 K-----SKEE---------------TNKTNEQVYRS----------INQNNTLILKENEI 1204

Qy        239 DLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKN-HIPFRDSELTKVLRD 297
Db       1205 D----ENDINTLNQNLNIKNDMNNMD-NIINLKN----LNNINNIYTPYQNNILKN--NE 1253

Qy        298 I-FVGKSKSIMIANISPTISCCEQTLNTLR-YSSRV---KNFKNKSTCINEED---DTNT 349
Db       1254 IQFLNNNKEV-ITKHAYTHSSNEININVYKNIDTQINIRKNENNNDIIINKEQKDISTNN 1312

Qy        350 ER-------------ISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKI 395
Db       1313 EQNQTYNYITSSVKNEYSILDNSPTDKNYKKNLNLNISSSY----NTNDHIN--NYHDDI 1366

Qy        396 ERNNILKNKSFDKPR--------EGFTSTFGK-YSSLNDIDKIKKNKKKGLINYKSTLYN 446
Db       1367 DKTN--KDNKYEEEKHKDINIYINNINSNGNKNHNFINSYFDLNENEKKKIN---IFYH 1421

Qy        447 DNTINKKHNNNNNNN-------NDNNDNNNDNNN--------NN---NND--------SS 480
Db       1422 DNKSLKNISNEELTNTNPYKCINNTIDNNNDHKNGYIELNEINNSIMNNDEKIITAYLNN 1481

Qy        481 SMVNNMINHMI----------------NNNINNNINVNNN---NNNNNNNNNSHNNHLPQ 521
Db       1482 RVVQNGLSHDILNEDVSINTHSNKENKNDNHNNNSNNSNNSSRSNNHNNNNNSNNNY--- 1538

Qy        522 PNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHL 581
Db       1539 -------ANSYNVKNDL--QINNNNYSYIKNEKNVHQTF------NENNILYK-----HY 1578

Qy        582 ARHSVGSKLTMFSYDPQKN--KDNTF------FKSNI-NKMEDNTP-KDILYESRNVSNM 631
Db       1579 SNH-VHNYITKNTIQEQKNIHSDNKYETCSLPEKKNIYEPNSQNSPNSQKLPNSQNLSNS 1637

Qy        632 NG-------------------NVLLGLNKNTHHDIST-----------KDENHNDNKIN 660
Db       1638 QNLPNSPNLPNSQNLPNSPNLPNCFTYVNNQTDILISNLVVQEQKLNDQKKEMRSNNNIN 1697

Qy        661 NGVINIINNSNVNSINNSNMN-----------SINNSNMNSNSIYKSNYNSNQSISDVQI 709
Db       1698 NTKWNNAYNNEVQHVNNNIVNEEINILPYPSKNENNLNEHNSNIYHDKNNQH----DIRT 1753
```

```
Qy     710 RYVNEMDTSNK-----NNDNIFFDAISCDN----------------------------- 734
Db    1754 YYNDSTKITNKLHIESGNNKLTNDAIHTNNEYTINSRYPICQNYSAITAPEKMNCPQMMA 1813

Qy     735 -------------NMYPNITNNNNNN--NNNNNNNNNNIDVEN--------YNNR--DGT 769
Db    1814 NHNFNNIPEYVIQNNHVNVDGNNKNDPIDNNNNNDNDNIEIFNIQQTKQIKYNNNTTDTN 1873

Qy     770 NNSMKLYAYNSHNLFQPDNNKNTS--NIQNINTNK---------NNQDGNVNY-SMNFCH 817
Db    1874 NNSL------IHNV---ENNYNTDGLDISNISYNNYTNNNMYNLNNDNINNNYMPQNYYH 1924

Qy     818 YNLN-----DKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNN 872
Db    1925 INYNNIATSENNQLIQINN-YNINTNIN--DININTHVNSYKENDKKNILYNNNI----N 1977

Qy     873 NMGNN--NSPRMKYGLCGSHTSIDNMKNNEMKNNEMKD-----NEMKDNHIKSNNNNSSS 925
Db    1978 NMNNDNINHDHYKY-ICNSNSNNNNNNNNNNNNNNNNNNDHYYYYNCCRNEEYKYNDNNYIN 2036

Qy     926 SSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIY----------QNDDII 975
Db    2037 IQNKKNNIVND-NTSNIIKSQY-NNTNKNMILEQHNINMNEDNYNQINSNLCLLKNDSIS 2094

Qy     976 YTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIY---------------------NKEMEG 1014
Db    2095 TCCDNINNNNNNN----NNKYSNNEINSNEYLYYKFDEQYDKKNNVSAMCDNIKNKQHYN 2150

Qy    1015 KHIRLDDQDKYDD-----------------------NDNNNVDNNN-------------- 1037
Db    2151 NYVAYNTNGSYENMSNFSSNNVYVNNKCPSYNNNNMNNMNNMNNNNCYNYHGYNVMNTKY 2210

Qy    1038 ------------------------------------------KNNVDNNVDNNNVDNNVDNN 1057
Db    2211 NKAGNKFSSVQIDTQMDNPIHVDKNKNDDILERNIKGSIINNNNDSNDDNNNDDNNNDNN 2270

Qy    1058 DKNNVDNNNVDNDDDDVDF-HNIKNFNNNEYLSYFQ--KNVDTIINNCLNSLDISSMYDD 1114
Db    2271 NNDNASNNN--NDTNNINYLMNIKNTNNFDSTNNFDSTNNFDS-----TNNFDSTNNFDS 2323

Qy    1115 TK--------------------EILNNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDKT 1154
Db    2324 TNNFNNRNNFNNRNNFNNRNNFEHINNINKIHHIDNENNLCRE--DENNHNTILTNVKNT 2381

Qy    1155 AQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFYAYSRL 1214
Db    2382 TNKEY--INICTDHLNLLKNYNLNNYINLLKINKLNDYYYNSIINKSQDGTCNTFLQNSL 2439

Qy    1215 EKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKLVQD 1274
Db    2440 NNQ-SNIIQNNNIYENNQRIQ-LYNYPSFETYNNNIINTWNQKNTNNSYNSYIPLSNVQC 2497

Qy    1275 NIKN---SMDHNNIH 1286
Db    2498 NWKNHIFNVTQNNAH 2512
```

RESULT 34
Q8IBV8

```
ID   Q8IBV8      PRELIMINARY;      PRT;   3569 AA.
AC   Q8IBV8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF07_0053.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA   Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844506; CAD50890.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   3569 AA;  415260 MW;  8C2BBF5BD2D93F28 CRC64;

  Query Match             11.7%;  Score 799.5;  DB 5;  Length 3569;
  Best Local Similarity   23.5%;  Pred. No. 2.3e-24;
  Matches  338;  Conservative  228;  Mismatches  441;  Indels  429;  Gaps   66;

Qy     149 IYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRV-LTKEELILKMIDGVLLRKIGVNS 207
Db      96 IMLNKKRNMHEKYKLNVKKNNKNKKFNIKDRNMKEMYLTYDASHSDISDGVSL----YNE 151

Qy     208 QNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQN-KQTQTDGANINRS 266
Db     152 SNDNK----------FFKMNMDTSMDKNDLI--------IDDVTKNVLLNNNNDDNNNKN 193

Qy     267 LLALKECIRAMD-----------SDKNHIPFRDSELTKVLRDIFVGKSK----------- 304
Db     194 SLFLSNCEHEKDIMEYTTSGQYYDETNNI---NNKNTNIINNI-DNKNKYDNSSSNNLDY 249

Qy     305 -----SIMIANISPTISCCEQTL----NTL------RYSSRVKNF-KNKSTCINEEDDTN 348
Db     250 CINKYDIIIED--PDMNCEMETMIEKDNILTNHTDENYIDLNKSFNKNIRSDINEINNLS 307

Qy     349 T----ERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKN- 403
Db     308 TINRVEYLTYIQSTSSEKDNINNDTLYADNKEMLYIPKNN----------INVNNMINNY 357

Qy     404 -KSFDKPREGF-----TSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTI---NKKH 454
Db     358 HTYIDNPKNNYLNLIENNTMSEHQLLYDDDIVNEDRIEG--SYTNEISSTNNLPGDNQSN 415

Qy     455 NNNNNNNNDNNDNNNDNNNNNNN-----DSSSMVNNMINHMINNNINNNINVNN------ 503
Db     416 NNNKSNNNDNNNNNNNNNNNNNNNNFYYYNNNSECSSLLNDNKDTFFHSNYTINNITNEVQ 475

Qy     504 -----NNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNN-----NDKSIFLHK 553
Db     476 DKSIYNNSNNMYENGLRNNLSNDINLTYTINSN-SSNDENICDVNNKLYDSNYLTLNVNP 534

Qy     554 KNLRDNIKLKNRSSCDN--------------IMNKKKNNLHLHARHSVGSKLTMFSYDPQK 599
Db     535 LNYMEN------NLCDNKMYELTDYYNNSFVINEKNKNNMNLYHHS------------SM 576

Qy     600 NKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKI 659
Db     577 NVNNLNITEDFKRDGDYNGSTTIYNTIENNNNNNN---NNNNNNNNNNSNVDDGNNCH-L 632

Qy     660 NNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSN--------QSISDVQI-- 709
Db     633 GHNIHNINNVYNINNIHNNNMNSINNNDRRSNSIASDEPKTNITNQKIGKNNIDKVKIIS 692

Qy     710 ------RYVNEMDT----SNKNNDN------------------------------------- 724
Db     693 NNNSCTNYVNDVNTDCPNNNSNHDKKIPFKNDDYDTEMEEMQVKKREFKNMILNNKSEFT 752

Qy     725 ------------IFFD-----------------------------AISCDNNMYPNITN 742
Db     753 TINENDELYKTVLYFEKKNKKILKMRRSLKRKRKSPKNEHLENDSYAMQDDNNINEGQKN 812

Qy     743 NNNNN--------------------------------NNNNNNNNNNIDVEN-------- 762
Db     813 ENHSNDILNNQVVNNSNDINDKNMIHLKVEQQEDADSNNNNNNNNNNLNVHNKYTDNDNN 872

Qy     763 -----YNNRDGTN-NSMKLYAYNSHN------LFQPDNNKNTSNIQNI--NTNKNNQDGN 808
Db     873 LLLVLLNQKDNSNGNNME----NNRNIENNAMLIIPKKKNNSTNVKGVKRNTKTRNQQSD 928

Qy     809 VNYSMNFCHYNLNDKNYLIDLNNKEQK----------DKNIH-----GCDNNIIQNRNDF 853
Db     929 TN---DMC---VNNDEICGD-NGKTQKGVNRKVGPGRKKNVNFKMEENADTNVNYEKINI 981
```

```
Qy        854 EKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKD 913

Db        982 NNNNNSNDNNNNSIDNNNNSIDNNNN-----------SIDN-NNNSIDNN----NNSIN 1024

Qy        914 NHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNIN----SNIY 969

Db       1025 NNNNSIDNNNIIISSNNNNISNTMNGRKKIKK----LKSVQTLNNKCNNNLDLSCINKIN 1080

Qy        970 QNDDIIYTI--------------NSLNDYMSNTLLHFKEKYTY-PTLSTNEDIYNKEME 1013

Db       1081 NNNQMDYTVTDYSSGNSAILIDKQNCETDYLENV------QYNHTPNISLNGLTYN---- 1130

Qy       1014 GKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDN------NNVDNNVDNNDKNNVDNNNV 1067

Db       1131 GVMVKCELSDGKMDNNNNNIITANCNIKEEPVENMTCVYENKKNNNNNSNNSNNINNINI 1190

Qy       1068 DNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCL---------NSLDISSMYDDTKEI 1118

Db       1191 CNSSNS---NNYSNNNSNHFT--FMSNHNIEHNECVGKNNNQTIDHSEEPNNVKGKGKGI 1245

Qy       1119 LNNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVD 1178

Db       1246 RRNLRGRKGKNKKDD---DYFYE-------EEIHD--ENVAKKRNKSKNKNTL--KRNKK 1291

Qy       1179 TQINNETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENL--RL 1236

Db       1292 CDDNNINSHMGIMITQNVQENNTGNITFN----DNINKSNTNYRFSNNIQVKNENINNNI 1347

Qy       1237 LYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKLVQDN------IKNSMDH-NNI 1285

Db       1348 LYNDISTICNNNNNNNINMNVINGNSENCITNEN--HDNHNTNKSISNDKDYLSNI 1401
```

```
RESULT 35
Q8I4X5
ID   Q8I4X5      PRELIMINARY;      PRT;   1605 AA.
AC   Q8I4X5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFL2190C.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014851; AAN36522.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1605 AA;  188160 MW;  CBDB7FA9E1D7C268 CRC64;

  Query Match             11.7%;  Score 798;  DB 5;  Length 1605;
  Best Local Similarity   22.4%;  Pred. No. 1.3e-24;
  Matches  384;  Conservative  222;  Mismatches  455;  Indels  652;  Gaps   75;

Qy         18 EKKKKDSDI---------ITVKNNCTLYIDE----PRYKVDMTKYIERH---------- 53

Db         50 EKKKKNEDIENNKAKYFEDVAINKNVEEYINVICRIKNNKIDINNKIDTNNKLDTNNKLD 109

Qy         54 ------------------------------------EFIVDKVFDDTVDNFTVYENTIKP 77

Db        110 TNNNINNNNVDVIKKISYNKLVIDTFHVGNKSSLNFEYTYDRVYDIENNNHMIFNDYIKN 169

Qy         78 LIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPY------------------------ 111

Db        170 NIKNIFQ-GINCSILAYGQTNSGKTYTMLGNFDFLNNLFHLCKDYNTCHNNENKINKVHD 228

Qy        112 ------------------------------------------------GQSDTPGIFQY 122

Db        229 KINHKDNNDNNKDATHNNNNIIKTNINMNNSDQHNDIHLTYDELTNCITNEPNNVGIIPH 288

Qy        123 AAGDIFTFLNIY---DKDNTK----GIFISFYEIYCGKLYDLL--QKRKMVAALENGKKE 173

Db        289 CINYIFNYINYHKQQDKINDKIKEFTVTLSILEIYNEVIYDLISGESNLSVHMIDSNKNE 348

Qy        174 VVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTS-- 231

Db        349 FIIKNLKEVEIENIISALYYLEQGVNNRKIAFTHMNKASSRSHLIFIIKINRYIYSTNTI 408

Qy        232 -LGKIAFIDLAGSERGADTVSQNKQTQTDGA------NINRSLLALKECIRAMDSDKNHI 284

Db        409 RCGKLCLVDLAGSER-------LKQTKATGSIKIETTMINKSLTVLSKVI---------- 451

Qy        285 PFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEE 344

Db        452 ------------------NSLAIMQIKEKIE--------------KNIKEKEK--NKD 475

Qy        345 DDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNK 404

Db        476 NDKD-------DDKGNDK---------------GNDKGNDKDNDKGNDK---------- 502

Qy        405 SFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDN 464

Db        503 -----------------GNDKD------------------NDKGNDKGNDKGNHNINDN 526

Qy        465 -NDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNN-NNNNNNSHNN----H 518

Db        527 INDNLNDNLKDNLKDN---LKDNLNDKDKYSCDNNININNNNSHINNSSHNPHSNTCNIH 583

Qy        519 LPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFL------------------------- 551

Db        584 IP-----YRDSKLTRVLSD---SLGNNCKSILICTLSSKLQYLNETASTIKFAQRAKMVK 635

Qy        552 -------------------HKKNLRDNIKLKNRSSCDN----------------IMNK 574

Db        636 AKPIVREEKVENKKEADTNEKEKNMKKKKKKRNEEKNDNNGDDYKEDNYDDDDDDDDDD 695

Qy        575 KKNNL--------------HLARHSV--GSKLTMFSY---------------------- 595

Db        696 EKNNLYDNKLNIFLNFNKNHITKDIIFFCFSLCIFSYICGFKSTGKVKYLDFFIESYKNK 755

Qy        596 ----DPQKNKDNTFFKSNINKMEDNTPKD--ILYESRNVSNMNGNV-----LLGLNKNTH 644

Db        756 IDIIKEQLNKSNESNKNDINEYLMKTFDDLKVLTEEERQRNMKLNMEKLKKCKDDNKNTI 815

Qy        645 HDISTKDEN---------------------------HNDNKI-NNGV--INIINNSN-- 671

Db        816 NQIQKKQENEEHALTIRNLAKVDPTLISKYDIQKKYQHNNNDVYDEGVHYLNVDDNNNDD 875

Qy        672 VNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNE----------MDTSNK 720

Db        876 VHADNNDETHIHDNDNDNDDDVL---------IGDILVKMKNDICNIIKFNLTSLKCNRK 926

Qy        721 NNDNIFFDAISCDNNMYPN--------ITNNNNN-----------------NNNNNNNN 754

Db        927 ENDEIFEEILHIVNSVKENNDIFLINELKNNSKDEKKINKKKDAQGDKINIPDDNNNNNN 986

Qy        755 NNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMN 814

Db        987 NNKISFKSIKNM----NIVKKYLKKD----DVNKNKNTKSILNINHKDDNK--------- 1029

Qy        815 FCHYNLNDK-NYLIDLNNKEQKDKNIHGCDNNIIQNRND------FEKK--KKTNFYNNN 865
```

```
Db   1030 -CYQDNNDNDNYITG------DDKNM----NNNIKNKNDILHYQVFAQKILSVTNIQNIF 1078

Qy    866 NIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNN---- 921

Db   1079 NYV-------NHN-----YTKFISDFTTD--KSTVFSNFDLKEN--LDRQIIEKNNTIYN 1122

Qy    922 ---------NSSSSSSSNNN---------IYNN------------INDDDTF----QNDY 947

Db   1123 FIHSATMFQHSQTLKNLNSNYFINKIQTEIYNNYPHIIHMKYMPSINNNEPIITERQNDL 1182

Qy    948 CHN---DNTFTIRRKNNTNINSNIY-------------QNDD---IIYTINSLNDYMSNT 988

Db   1183 IVTTPLKNILQINKEYNNTYNTHKYIPIKDIIPDHTTNQNETTCAINKSVTFFNEHGGNK 1242

Qy    989 LLHFKE----KYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNN 1044

Db   1243 --HSEEDLSNKNNIPIEENKKEINNKSITFLRQSLEGTKNKQDGSFSNIAPTNKNN-KNY 1299

Qy   1045 VDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEY--LSY----------- 1090

Db   1300 DDDNNKNNKYDDDDNNNNNNNNINNDDDNICVGEKKHIRNHAHKDQSYEVQNQKDKNSKK 1359

Qy   1091 ---FQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNV--IKKYI-NEDIK 1144

Db   1360 KGSFVHNIN-VSDKSDNSLEFTKLLED-KDFKK---LYEYLIEKENTNDVNTFIQKEEHM 1414

Qy   1145 NMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINN---------ETSDLRKDLVMC 1195

Db   1415 NNALNELECKINS--EIRRVSDHIVVDYSKNNHKNKHNNMNHNENSSRKNSNINKN---- 1468

Qy   1196 HICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVE--YQNKS-ANS 1252

Db   1469 DIYNNN--DNMIMVVKKKKEKNYNNLDISYLEKKRNTNLFIKKIFHIIKNIFQKKDFKEP 1526

Qy   1253 VLLNVSSNNGDIILLNKKLVQDNIKNSMDHNNI 1285

Db   1527 IEINVSDIIENELLHNKKLLEKN----LDSYNI 1555

RESULT 36
Q8IL84
ID   Q8IL84      PRELIMINARY;      PRT;  1922 AA.
AC   Q8IL84;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Metacaspase-like protein.
GN   PF14_0363.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014822; AAN36976.1; -.
SQ   SEQUENCE   1922 AA;  226572 MW;  5E72815A3662B31F CRC64;

  Query Match             11.7%;  Score 797.5;  DB 5;  Length 1922;
  Best Local Similarity   24.6%;  Pred. No. 1.6e-24;
  Matches  365;  Conservative  221;  Mismatches  498;  Indels  399;  Gaps   72;

Qy      1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 60

Db    171 INGKERLLNR----YNIMRPQRSADML-IRN--YLYNDLP-YEINQQN--------LNKN 214

Qy     61 FDD----TVDNF--TVYENTIKPLIIDLYENGCV----CSCFAYGQTGSGKTYTMLGSQP 110

Db    215 INDNNAGSFDNLNKTVFENYNKN-ILDEYNRNIMQYYNSNLKNYRNSSFDNIYTRYNNNN 273

Qy    111 YGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENG 170

Db    274 NNNN------VYSNNNVYSNNNIYNNNNNNDNINNSFQF--SKQEGLM---KIYSNNNRN 322

Qy    171 KKEVVVKDLKILR----VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDI 226

Db    323 MNDYVVKTSNITRAADNTCAQEQINIPTYNKTVLLK---SSPNTENTFNKTLSNI-LKEQ 378

Qy    227 NKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPF 286

Db    379 PKRSSM-----------------------------------------------FSF 387

Qy    287 RDSELTK---VLRDIFVGKSKS----------------IMIANISPTISCCEQTLNTLR- 326

Db    388 NSSDFFKDFGKLRKNFVGKKKNKINKDIVINGNIFNNKIIDNNTSNAKSEGNSTLHKIYN 447

Qy    327 --YSSRVKNFKNKSTCIN--EEDDTNTERISILDSKGS--EMNASSIENVVIKSNHLLSN 380

Db    448 FIFPSSVKKLQKKKQKLQVIEDKEKNSNRAVVMPLKNKIIYINDANITN--DKFKHSFEN 505

Qy    381 NNNNKINRGKIN-----DKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKK 435

Db    506 YNYHQTNIPFVSVTNYMNKDSTKNVRSFSSFD--------------------VKERISP 544

Qy    436 GLINYKSTLYNDNTINKKHNN-NNNNNNDNDNNNDNNNNNNNNDSSSMVNNMINHMINNN 494

Db    545 RVIR------NGNINNNKNNNINNNNNNINNNNNNINNNNNNINNNNNINNNNNINNNNN 598

Qy    495 INNNINVNNNNNNNNNNNNS--------------HNNHLPQPNYAFTDTSDFSSLDDMNC 540

Db    599 INNNNNINNNNINNINNINNNDFVNDSNINKTVNIHNNNNNLNNYSNTMKSVAKNTQKLSA 658

Qy    541 HLNN--NDKSIFLHKK---NLRDNIKLKNRSSCD----NIMNKKKNNLHLARHSVGSKLT 591

Db    659 QKNSLVNDKHHTYHEKLITNIYQFTGGQSQKSCNIKNQNVKEEKINNTQI------SYIN 712

Qy    592 MFSYDPQKNKDN--TFFKS----NINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHH 645

Db    713 NTAYNDYTNLPNVETYVLSEKSNNLENKEDNNINEKFKNDNNVDKKN------------- 759

Qy    646 DISTKDEN----HNDNKINNGVINIINNSNVNSINNSN----MNSINNSNMNSNSIYKS- 696

Db    760 ----NDENNIIKHNVNQNNNEVI-VMNNQTEEQINDNKFIKPLNNITGIDPKKKNIYNTV 814

Qy    697 --------------NYNSNQSISDVQIRYVNEMDTS-------NKNNDNIFFDAISCDNN 735

Db    815 MHINDKMDLNKHDYNYNYNNIIPH---NYINSSDKMTNNSLHINKKADDIFQNMISNNDT 871

Qy    736 MYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNI 795

Db    872 NVPCVYVNKYINNIFQNNNPSN--VTNYINTINMNKAPITIDENSNKL---DNNINYENV 926

Qy    796 QNINTNKNNQDGNVNY--------SMNFCHYNLNDKNYLIDLNNK----EQKDKNIHGCD 843

Db    927 TKNMIYNNYTNNNVLYITKEKSDNKLN------NSTNHLNDLKNNYYVINPSLLYIHNNN 980

Qy    844 NNIIQNR---NDFEK---------KKKTN----------FYNNNN-------IVIVNNNM 874

Db    981 NNIEWNKQYTNNFSKPNFLDKFFYEKKVESIQPHHLNQYYYNNQNVKYNPSHINILENKY 1040

Qy    875 GNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIY 934

Db   1041 VNQFTNQNENSLYNFTTHHNNIKNNNIDNILLLDKAIRNQSIYYNNNKGDDETKSAEHTT 1100
```

```
Qy        935 N-NINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFK 993
Db       1101 NPSIN---LLKNNY---------EDKNKSN----------YVHSIVSITD--KDNSIYAK 1136

Qy        994 EKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDND-------NNNVDNNNKNNVDNNVD 1046
Db       1137 NKYAEVVTKNNE----KREDAIH-KKDEQYIFNNNDNNNDNIHNNNIHNNIHNNIHNNIH 1191

Qy       1047 N---NNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIIN--- 1100
Db       1192 NNIHNNIHNNIHNNIHNNIHNNIHNNIHNNIDVSSRENFFQHDIQNTNQINNKTILDITQ 1251

Qy       1101 ----NCLNS--LDIS----------SMYDDTKEIL----------------NNILLSKY 1127
Db       1252 NEKPNATNEKCLDINNKEKKKKKIKSVHYDMDEILSLKCKVYDELDTCLKCINNLTKNKK 1311

Qy       1128 KAEKDNVIKKY---INEDIKNMSLEE---IDKTAQSIYEKRKVLLTKLLL-LFKKN-VDT 1179
Db       1312 LYEQ---LKKYKHKLDKDTKTKKLYDECTSEKTG-SYYTMKDSYISSLLCNEYKLNHILN 1367

Qy       1180 QINNETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQ 1239
Db       1368 KSNNKVSDDK------HILSYN-----HHHNKNNNINDINNIKHISK-FVNKDNEKTLNN 1415

Qy       1240 FLVVEYQNKSANSVLLNVSSNNGDIIL---LNKKLVQDNIKNS 1279
Db       1416 INILNHNNKDPNSGVFLPSDNTESVIYNMNRNRKLCDNNFNVS 1458
```

```
RESULT 37
Q94GW1
ID   Q94GW1      PRELIMINARY;      PRT;   800 AA.
AC   Q94GW1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Kinesin-like protein.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hsing Y.-I.C., Chow T.-Y., Wu H.-P., Chao Y.-T., Liu S.-M., Lee P.-F.,
RA   Su C.-L., Chen C.-S., Shaw J.-F.;
RT   "Oryza sativa PAC P0431G05 genomics sequence.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC087551; AAK70904.1; -.
DR   Gramene; Q94GW1; -.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   ATP-binding; Coiled coil; Microtubules; Motor protein.
SQ   SEQUENCE   800 AA;  88189 MW;  D9B330375AC6F256 CRC64;

  Query Match             11.6%;  Score 796;  DB 10;  Length 800;
  Best Local Similarity   34.7%;  Pred. No. 8.2e-25;
  Matches  207;  Conservative   96;  Mismatches  201;  Indels   92;  Gaps   18;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFD 62
Db        198 AKIKVVVRKRPLNRKEVSRKEEDIITVHDSSSLTVYEPKLKVDLTAYVEKHEFCFDAVLD 257

Qy         63 DTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY 122
Db        258 EQVSNDEVYRETVEPIIPIIFQR-TKATCFAYGQTGSGKTYTM---QP--------LPLR 305

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
Db        306 AAQDMVRLLHQPVYRNQNFK-LWLSYFEIYGGKLFDLLSDRRQLLMREDGKKQVCIVGLQ 364

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
Db        365 EFEV-SDVQIVKEYIERGNAARSTGSTGANEESSRSHAILQLAIK--KHIIAVGKISFID 421

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
Db        422 LAGSERGADTTDNDRQTRIEGAEINKSLLALKECIRALDNDQIHIPFRGSKLTEVLRDSF 481

Qy        300 VGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKG 359
Db        482 VGNSRTVMISCISPNAGSCEHTLNTLRYADRVKSL-SKGSNTRKEQPTGPTIPSSKDSSS 540

Qy        360 SEMNASSIENVVIKSNHLLSNNNNNKINRGKINDK--IERNNILKNKSFDKPREGFTSTF 417
Db        541 APSYPMPIETEEIAN--------------QIQEKRPVETSR--------KAAENFTSNS 577

Qy        418 GKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNN--DNNNDNNNNN 475
Db        578 SMEPDRNPVSMIPSYSNRG---------------KEENGSSGLNDRERVDLNSSRISYN 621

Qy        476 NNDSS--SMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQ---PNYAFTDTS 530
Db        622 SKPQSVQSSANLQEEEKVTKVSPPRRKAYRDDKPERQSNYAKKDSGPETSRPGYKVQQAK 681

Qy        531 DF----------------------SSLDD--MNCHLNNNDKSIFLHKKNLRDNIKL 562
Db        682 QLQQQQRPTSASASQNSSRQSEKESSCDDVEIDAILEEEEALIAAHRKEIENTMEI 737
```

```
RESULT 38
Q9LW81
ID   Q9LW81      PRELIMINARY;      PRT;   706 AA.
AC   Q9LW81;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similarity to kinesin heavy chain.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RA   Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RX   MEDLINE=20277480; PubMed=10819329;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT   features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:131-135(2000).
DR   EMBL; AB012247; BAB02671.1; -.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   ATP-binding; Coiled coil; Microtubules; Motor protein.
SQ   SEQUENCE   706 AA;  79246 MW;  059658E9E69282F2 CRC64;

  Query Match             11.6%;  Score 795.5;  DB 10;  Length 706;
  Best Local Similarity   36.8%;  Pred. No. 7.6e-25;
  Matches  203;  Conservative   91;  Mismatches  186;  Indels   71;  Gaps   16;
```

```
Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFD 62
Db        168 AKIKVVVRKRPLNKKESTKNEEDIVDTHANC-LTVHETKLKVDLTAYVEKHEFVFDAVLD 226

Qy         63 DTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY 122
Db        227 EEVSNDEVYRETVEPVVPLIFQR-IKATCFAYGQTGSGKTYTM---KPLPLK-------- 274

Qy        123 AAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKIL 182
Db        275 ASRDILRLMHHTYRNQGFQLFVSFFEIYGGKLYDLLSERKKLCMREDGKQQVCIVGLQEY 334

Qy        183 RVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDL-KDINKNTS-----LGKI 235
Db        335 RV-SDTDAIMELIERGSATRSTGTTGANEESSRSHAILQLAIKKSVEGNQSKPPRLVGKL 393

Qy        236 AFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVL 295
Db        394 SFIDLAGSERGADTTDNDKQTRLEGAEINKSLLALKECIRALDNDQGHIPFRGSKLTEVL 453

Qy        296 RDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFK----------------NKST 339
Db        454 RDSFMGNSRTVMISCISPSSGSCEHTLNTLRYADRVKSLSKGNASKKDVSSSTMNLREST 513

Qy        340 CI----------NEEDDTNTERISILDSKGSEMNASSIE--NVVIKSNHLLSNNNNNKIN 387
Db        514 KIPLSSALPTPSNFDDDVN----EMWTEENDEFDASDYEQDKQMWKKNGKLEPSYN---- 565

Qy        388 RGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYND 447
Db        566 -GMAQERIPKPTI-QMKSRDMPRPDM-----KKSNSDDNLNALLQEEEDLVNAHRKQVED 618

Qy        448 NTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSS-------MVNNMINHMINNNINNNIN 500
Db        619 TMNIVKEEMNLLVEADQPGNQLDGYISRLNTILSQKAAGILQLQNRLAHFQKRLREHNVL 678

Qy        501 VNNNNNNNNN 511
Db        679 VSTTGERKSQS 689
```

RESULT 39
Q8IEI5

```
ID   Q8IEI5      PRELIMINARY;      PRT;  1192 AA.
AC   Q8IEI5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Plasmodium falciparum asparagine-rich protein.
GN   MAL13P1.63.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844509; CAD52272.1; -.
SQ   SEQUENCE   1192 AA;  139166 MW;  9CEE8FDDBD0A5A46 CRC64;

  Query Match             11.6%;  Score 795.5;  DB 5;  Length 1192;
  Best Local Similarity   25.5%;  Pred. No. 1.2e-24;
  Matches  295;  Conservative  163;  Mismatches  367;  Indels  332;  Gaps   48;

Qy        327 YSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLS--NNNNN 384
Db         27 FRNSIKNNENLIFPANNKKD------------------NGMEQECWRSNKKFDDMNNRNN 68

Qy        385 KINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTL 444
Db         69 SLNNNDMNNNNNNNNM--NNS-DNSKLSFSRNF-----------KNSFNKGMNRNNSVT 113

Qy        445 YNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNN- 503
Db        114 GNIKGINNKNYNYMNNNQENGTNNNDMNVNDQNNYSGIFNNNFTNLNNKAFYKNMNKNNK 173

Qy        504 -------NNNNNNNNNSHNNHLPQPNYAFTDTSDF-SSLDDMNC--HLNNNDK------ 547
Db        174 FNKTNNRNNNNNNNNNKNNNNNDDGNINYQNTNEFKDNKKNMNFKNQYNNNYKFDENMN 233

Qy        548 -SIFLHKKN------LRDNIKLKNRSSCDNIMNKK---------------------KN 577
Db        234 NSNTMHSRNSNVEEHLRNNSIDMNNSNINNYTNQQTRFSSFMENENENENKNYHTGGMNN 293

Qy        578 NLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSN------- 630
Db        294 NIHFKNKYDNNNSSMKNTDNNKTDTSYNMKGTIN--NDNNNMDYL---RNINNINEYKGS 348

Qy        631 ---------MNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNM- 680
Db        349 AKNKFYTNYMNKNNLKFTQNNNDNMNTNEDNNNNNNNNNNGVFSNYQNNNMNRNNSINIK 408

Qy        681 -NSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKN---NDNIFFDAISCDNNM 736
Db        409 RNLNNNNNINNN--------------------MNKMGSQDKNQNSNNNFYM------NYN 442

Qy        737 YPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQ 796
Db        443 YQNRKNSMNNNMNNNMNNNMNHNMNNNMNHNMNNNM-------NHNM---NNNMN----H 488

Qy        797 NINTNKNN-----QDGNVNY----SMNFCHYNLNDKN-------YLIDLNN----KEQKD 836
Db        489 NMNNNMNNINSLDSDMSPNYHAHVKMSMMNYNNNESNTANPNQMNFEQTNNDNMKRENNN 548

Qy        837 KNIHGCDNNIIQNRNDFEKKKKTNFYN-----NNNIVIVNNNMG---NNNSPRMKYGLCG 888
Db        549 MNNYGYDDNTVHVNNN---TPSTDFFSRAVGYNNNYLNNNNNMNSAVNNNS--------- 596

Qy        889 SHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSNNNIYNNINDDDTFQNDYC 948
Db        597 --SNGNNMKNENSENKNVADNNDSLNNNKNNNNNINMNESINNN--NTLNNN----NEY- 647

Qy        949 HNDNTFTIRRKNNTNINSNIYQND------DIIYTINSLNDYMSNTLLHFKEKYTYPTLS 1002
Db        648 -----------NNQNNNEDEDDDDWGELGEDKYIDINSIMK-KKNVILNQLE-------A 688

Qy       1003 TNEDIYNKEMEGKHIR------------------LDDQDKYDDNDNNNVDNNNKNNVDNN 1044
Db        689 DLNDLSKKGNDGKNKKKNKMKKDDLFVLPHTNTLLVDKKKNKKNKNKNAKNNNTNSNNNN 748

Qy       1045 VDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNE----YLSYFQKNV----- 1095
Db        749 NNNNNNNNNNNNNSNILKKDNDANDKNSKQIVNASGVNKKKGKKTGVEANSKNVEQKGD 808

Qy       1096 -------------DTIIN-----------NCLNSLDISSMYDDTKEILNNILLSKYKAEK 1131
Db        809 TKTSTKVDEEVKTDEVNNKKKDLIEDKNANNVNDVDTKPDGDDDNNNNN-----NKGEE 863

Qy       1132 DNVIKK--------------YINEDIKNMS-LEEIDKTAQSIYEKRKV-----LLTKLLL 1171
Db        864 DNKEKKEDDNTVAEKPAKALYVFKKKENMSPIEYMERQVKSLLNKLTVENFPIITEKMCQ 923

Qy       1172 LFKKNVDT-QINNETSDLRKDLVMCHICNNNPDDQFHFYA------------YSRLEKDI 1218
Db        924 IMESRLNTDEIQTVVNQVIDKAVLEH-------DWSEMYADLCQTLKWRSPNFEMKKKTS 976

Qy       1219 INLIMLRQIWCESENLRLLYQFLVVEYQNKSANSV------------LLNVSSNNGDIIL 1266
Db        977 FEIALLKKIQEQYENLPSTFESTMKEKLKSDENEEELSFVEQKQKKRLLGIVKLIGE--L 1034

Qy       1267 LNKKLVQDNIKNSMDHN 1283
Db       1035 FQRQIVSISIVISIAHD 1051
```

```
RESULT 40
Q8ICZ4
ID   Q8ICZ4      PRELIMINARY;      PRT;  1437 AA.
AC   Q8ICZ4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   MAL6P1.23.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844505; CAD50294.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1437 AA;  165399 MW;  F0ADA05528A741DA CRC64;

  Query Match             11.6%;  Score 795;  DB 5;  Length 1437;
  Best Local Similarity   24.8%;  Pred. No. 1.5e-24;
  Matches  311;  Conservative  195;  Mismatches  405;  Indels  342;  Gaps   56;

Qy        219 LNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDG-ANINRSLLAL------- 270
Db        308 MNFDILD-NNAYKVVEFCFTDYENNK------NKNNCNWKDGNENFNNSSMTSYFSNNGK 360

Qy        271 -----KECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTL 325
Db        361 ILENEKSNIYHSNNNYNNNNNNDNNQTSTHFDEALNYNKT-----------CEYLPNLL 408

Qy        326 RYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNK 385
Db        409 T-SEEVNDNNNKCKKFNTLPTSFKDNIA-----GNIMNNSNNSCDMLNSKYVLNNSNNTS 462

Qy        386 I------------NRGKIND--KIERNNILKNKSFD----------------KPREGF 413
Db        463 IFDNYTNLSTLFDMNDSKMNNLTNFSNCNTLEHLSFDELNKSNNNNDDFFISNEKSEDGK 522

Qy        414 TSTFGKYSSLNDIDKIK---KNKKKGLINYKSTLYNDNTINKK------HNNNNNNNDN 464
Db        523 NLKFSQFLSSNSSNGVRYGANNNDLSIVNMCGLDTNDN--NKKEDNPMDNNNNNNNNDD 580

Qy        465 NDNNNDNNNNN-------------NDSSSMVNNMINHMINNNINNNINVNNNNNNNN 510
Db        581 NINNNNNNNNNNFLNKEESKEAKTEGTNSNNIQSNENNTKPGNNKNSKKNKNKKNALNNI 640

Qy        511 NNNSHNNHLPQPNYAFTDTSDFSSL--DDMNCHLNNNDKSIFLHKKNLRDNIKLK----N 564
Db        641 RNNSINRNLKGVNF-----SEVVSLNNNEMKSEVQEKKKNNKKENKNNKNNCKEKDQEEN 695

Qy        565 RSSCDNIMNKKKNN--------------LHLARHSVGSKLTMFSYDPQKNKDNTFFKSN 609
Db        696 KISNEKHINKKEKNKNAGEDDEEDVTITDINVLEKEKSKNQKIKNNENQHNKKNNIGQNN 755

Qy        610 INKMEDNTPKDILYESRN---VSNMNGNVL------LGLNKNTHHDISTKDENHNDNKIN 660
Db        756 INNNNNNN------NSKNGGRNSKKGGDIINNQQTVAGNNKKSEIKNGEKNEKSNNHNNN 809

Qy        661 NGVINIINN----------SNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIR 710
Db        810 DNLFNNKNNCSYTQMNKKFTDVNVIENNHISDVNHAK-SSISVQSKNNENNMN------- 861

Qy        711 YVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNNRDGTN 770
Db        862 --KEKDNKKKNNNN---------NN------NNNNNNNNNNNNNNN----NNNNNNNN 900

Qy        771 NSMKLYAYNSHNLFQPDNNKNTSNIQNI-------NTNKNNQDGNVNYSMNFCHYNLNDK 823
Db        901 NILNNISNKENKVI--TNQKNNGLVGNMSGTIISTNHSNNNERGNKKY-MNNYHVNSH-- 955

Qy        824 NYLIDLNNKEQKDK-----NIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNN 878
Db        956 -----VHNKENKNNINNRINVISNNNN---NINNIEEDSK----NNSNDILTNLNDTNIH 1003

Qy        879 SPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNN---------------- 921
Db       1004 QLLMNHGQLDDK---DNMKNNKKDDNNIAFNVLTNKENKCNNNYNNIGRAITTAVGTSSF 1060

Qy        922 -NSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINS 980
Db       1061 INNSNIGTENKNFVNDIIGD---------------IHEKGRDN-NSIFLETDN----YNN 1100

Qy        981 LNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGK------HIRLDDQDKYDDNDNNNVD 1034
Db       1101 LNKLI---------------MQENENMFNKNIRSNVKNILDYITMENKNMNNNNNNNNN 1145

Qy       1035 NNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKN 1094
Db       1146 NNNNNNSNISDDNALFSNNNYMNKHENLYTNNMINGNKIVDFDNMRNMNSN----YNKK- 1200

Qy       1095 VDTIINNCL----NSLDISSMYD----DTKEILNNILLSKYKAEKDNVIKKYINEDIKNM 1146
Db       1201 -----NDCMFLFKDNLSISGLGNVNLCNEKNMINESLNNIFKGDNIN----------NNM 1245

Qy       1147 SL---------EEIDKTA--QSIYEKRKVLLTKLLLLFKKNV-DTQINNE---TSDLRKD 1191
Db       1246 SLGINSPCLTKEDFFNCAGNNNVVENKSK---------EYNISDNGINDSCDYTNNLLNI 1296

Qy       1192 LVMCHICNNNPDDQFHF-----------YAYSRLEKDIINLIMLRQIWCESENLRLL-- 1237
Db       1297 FLSCNSINENFDDANEYNNKINGFKGENHYSYNNVGKNISTDI--NESYNNSENHFNLFG 1354

Qy       1238 -YQFL------VVEYQNKSANSVLLNVSSNNGDIILLNKKLVQDNIKNSMDHN 1283
Db       1355 NYNFFGNNCIHCKSYKNEIKN-LTLQIKLMKEEILKYKQLIMSSGINEKFEKN 1406

Search completed: October  2, 2003, 16:20:40
Job time : 184 secs
```

OM protein - protein search, using sw model

Run on: October 2, 2003, 16:28:21 ; Search time 131 Seconds
(without alignments)
2537.189 Million cell updates/sec

Title: US-10-006-780-2
Perfect score: 1288
Sequence: 1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ND
Not a [illegible]

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1288	100.0	1351	5	Q8I4Y0	Q8i4y0 plasmodium
2	19	1.5	1723	5	Q8T224	Q8t224 dictyosteli
3	18	1.4	348	5	Q8MN71	Q8mn71 dictyosteli
4	18	1.4	439	5	Q8T832	Q8t832 dictyosteli
5	18	1.4	699	5	Q8MN10	Q8mn10 dictyosteli
6	18	1.4	815	5	Q8T686	Q8t686 dictyosteli
7	18	1.4	1518	5	Q8T1J6	Q8t1j6 dictyosteli
8	18	1.4	1887	5	Q8SSS8	Q8sss8 dictyosteli
9	18	1.4	2994	5	Q95ZG5	Q95zg5 dictyosteli
10	18	1.4	3177	5	Q9NA13	Q9na13 dictyosteli
11	18	1.4	3619	5	Q9U573	Q9u573 dictyosteli
12	17	1.3	104	5	Q8MNQ5	Q8mnq5 dictyosteli
13	17	1.3	172	5	Q8MP24	Q8mp24 dictyosteli
14	17	1.3	204	5	Q8ST26	Q8st26 dictyosteli
15	17	1.3	279	5	Q8T1B5	Q8t1b5 dictyosteli
16	17	1.3	284	5	Q8ST23	Q8st23 dictyosteli
17	17	1.3	288	5	Q9XY02	Q9xy02 dugesia jap
18	17	1.3	293	5	Q8MX75	Q8mx75 cryptospori
19	17	1.3	300	5	Q8MNN2	Q8mnn2 dictyosteli
20	17	1.3	371	5	Q8ICP1	Q8icp1 plasmodium
21	17	1.3	384	5	Q8IHR1	Q8ihr1 plasmodium
22	17	1.3	391	5	Q25768	Q25768 plasmodium
23	17	1.3	419	5	Q25774	Q25774 plasmodium
24	17	1.3	426	5	Q8MNF6	Q8mnf6 dictyosteli
25	17	1.3	434	5	Q8T834	Q8t834 dictyosteli
26	17	1.3	440	5	Q8I4W5	Q8i4w5 plasmodium
27	17	1.3	455	5	Q8MXW8	Q8mxw8 dugesia jap
28	17	1.3	465	5	Q8IIJ6	Q8iij6 plasmodium
29	17	1.3	496	5	Q8SSX7	Q8ssx7 dictyosteli
30	17	1.3	498	5	Q8MP37	Q8mp37 dictyosteli
31	17	1.3	511	5	Q8SSZ5	Q8ssz5 dictyosteli
32	17	1.3	537	5	Q8T1K7	Q8t1k7 dictyosteli
33	17	1.3	560	5	O61085	O61085 dictyosteli
34	17	1.3	587	5	Q8MNG3	Q8mng3 dictyosteli
35	17	1.3	597	5	Q8MYG9	Q8myg9 dictyosteli
36	17	1.3	608	5	Q94474	Q94474 dictyosteli
37	17	1.3	614	5	Q8T1X4	Q8t1x4 dictyosteli
38	17	1.3	616	5	Q8MNJ9	Q8mnj9 dictyosteli
39	17	1.3	628	5	Q8SSY0	Q8ssy0 dictyosteli
40	17	1.3	634	5	Q8MN19	Q8mn19 dictyosteli
41	17	1.3	635	5	Q8MYH5	Q8myh5 dictyosteli
42	17	1.3	638	5	Q8T2T4	Q8t2t4 dictyosteli
43	17	1.3	647	5	Q8MMY6	Q8mmy6 dictyosteli
44	17	1.3	657	5	Q8MMU1	Q8mmu1 dictyosteli
45	17	1.3	662	5	Q8T1P5	Q8t1p5 dictyosteli
46	17	1.3	663	5	Q8MN61	Q8mn61 dictyosteli
47	17	1.3	682	5	Q9GPR8	Q9gpr8 dictyosteli
48	17	1.3	683	5	Q8MXN1	Q8mxn1 dictyosteli
49	17	1.3	723	5	Q8T158	Q8t158 dictyosteli
50	17	1.3	726	5	Q8MXL5	Q8mxl5 dictyosteli
51	17	1.3	728	5	Q9GYC7	Q9gyc7 leishmania
52	17	1.3	734	5	Q8T807	Q8t807 dictyosteli
53	17	1.3	740	5	Q8T138	Q8t138 dictyosteli
54	17	1.3	749	5	Q02602	Q02602 plasmodium
55	17	1.3	754	5	Q8MMT9	Q8mmt9 dictyosteli
56	17	1.3	756	5	Q8MYD6	Q8myd6 dictyosteli
57	17	1.3	759	5	Q9U478	Q9u478 dictyosteli
58	17	1.3	762	5	Q8MNN8	Q8mnn8 dictyosteli
59	17	1.3	791	5	Q8T2U7	Q8t2u7 dictyosteli
60	17	1.3	792	5	Q8I535	Q8i535 plasmodium
61	17	1.3	806	5	Q8T2A3	Q8t2a3 dictyosteli
62	17	1.3	821	5	Q8MNC3	Q8mnc3 dictyosteli
63	17	1.3	823	5	Q8MYH7	Q8myh7 dictyosteli
64	17	1.3	827	5	Q8MP44	Q8mp44 dictyosteli
65	17	1.3	828	5	Q8MN35	Q8mn35 dictyosteli
66	17	1.3	833	5	Q8SSX5	Q8ssx5 dictyosteli
67	17	1.3	863	5	Q8IDK7	Q8idk7 plasmodium
68	17	1.3	868	5	Q8MNP7	Q8mnp7 dictyosteli
69	17	1.3	887	5	Q8T1U0	Q8t1u0 dictyosteli
70	17	1.3	887	5	Q8I8F4	Q8i8f4 dictyosteli
71	17	1.3	917	5	Q8MMV3	Q8mmv3 dictyosteli
72	17	1.3	929	5	Q9BLX2	Q9blx2 dictyosteli
73	17	1.3	930	5	Q8T2J1	Q8t2j1 dictyosteli
74	17	1.3	931	5	Q8T2K1	Q8t2k1 dictyosteli
75	17	1.3	942	5	O96611	O96611 dictyosteli
76	17	1.3	946	5	Q8IC68	Q8ic68 plasmodium
77	17	1.3	959	5	Q8T1P2	Q8t1p2 dictyosteli
78	17	1.3	980	5	Q8MYI3	Q8myi3 dictyosteli
79	17	1.3	1041	5	Q8MMT7	Q8mmt7 dictyosteli
80	17	1.3	1057	5	Q8II33	Q8ii33 plasmodium
81	17	1.3	1060	5	Q8SSQ5	Q8ssq5 dictyosteli
82	17	1.3	1125	5	Q8MMQ6	Q8mmq6 dictyosteli
83	17	1.3	1130	5	Q8IJZ4	Q8ijz4 plasmodium
84	17	1.3	1139	5	Q23865	Q23865 dictyosteli
85	17	1.3	1197	5	Q8MP87	Q8mp87 dictyosteli
86	17	1.3	1224	5	Q8IAX9	Q8iax9 plasmodium
87	17	1.3	1231	5	O97140	O97140 dictyosteli
88	17	1.3	1297	5	Q8ICY8	Q8icy8 plasmodium
89	17	1.3	1309	5	Q8T2H9	Q8t2h9 dictyosteli

90	17	1.3	1321	5	Q8IAR1	Q8iar1 plasmodium
91	17	1.3	1338	5	Q23927	Q23927 dictyosteli
92	17	1.3	1363	5	Q8ICN3	Q8icn3 plasmodium
93	17	1.3	1376	5	Q8T235	Q8t235 dictyosteli
94	17	1.3	1407	5	Q8T1T5	Q8t1t5 dictyosteli
95	17	1.3	1422	5	Q8MMY3	Q8mmy3 dictyosteli
96	17	1.3	1437	5	Q8ICZ4	Q8icz4 plasmodium
97	17	1.3	1457	5	O44011	O44011 dictyosteli
98	17	1.3	1469	5	Q8IKP6	Q8ikp6 plasmodium
99	17	1.3	1477	5	Q8ILJ1	Q8ilj1 plasmodium
100	17	1.3	1483	5	Q8T273	Q8t273 dictyosteli
101	17	1.3	1485	5	Q8MNB3	Q8mnb3 dictyosteli
102	17	1.3	1502	5	Q8IS10	Q8is10 dictyosteli
103	17	1.3	1523	5	Q8I3B1	Q8i3b1 plasmodium
104	17	1.3	1543	5	Q9GV71	Q9gv71 dictyosteli
105	17	1.3	1546	5	Q95PI1	Q95pi1 dictyosteli
106	17	1.3	1548	5	Q95PI0	Q95pi0 dictyosteli
107	17	1.3	1565	5	Q8T1M2	Q8t1m2 dictyosteli
108	17	1.3	1570	5	Q9U0H8	Q9u0h8 plasmodium
109	17	1.3	1631	5	Q8I3H3	Q8i3h3 plasmodium
110	17	1.3	1670	5	Q23901	Q23901 dictyosteli
111	17	1.3	1672	5	Q8T2M6	Q8t2m6 dictyosteli
112	17	1.3	1709	5	Q95PH5	Q95ph5 dictyosteli
113	17	1.3	1721	5	Q8SSQ0	Q8ssq0 dictyosteli
114	17	1.3	1722	5	Q8II18	Q8ii18 plasmodium
115	17	1.3	1728	5	Q8SSU4	Q8ssu4 dictyosteli
116	17	1.3	1734	5	Q8MNM2	Q8mnm2 dictyosteli
117	17	1.3	1736	5	Q95PH7	Q95ph7 dictyosteli
118	17	1.3	1758	5	Q8ICB2	Q8icb2 plasmodium
119	17	1.3	1785	5	Q8I5Z0	Q8i5z0 plasmodium
120	17	1.3	1789	5	Q8T2E3	Q8t2e3 dictyosteli
121	17	1.3	1800	5	Q8I7P5	Q8i7p5 dictyosteli
122	17	1.3	1812	5	Q8IIS3	Q8iis3 plasmodium
123	17	1.3	1844	5	Q8IE31	Q8ie31 plasmodium
124	17	1.3	1880	5	Q8MP27	Q8mp27 dictyosteli
125	17	1.3	1894	5	Q8ISI8	Q8isi8 plasmodium
126	17	1.3	1894	5	Q8I3S3	Q8i3s3 plasmodium
127	17	1.3	1918	5	Q8SSW3	Q8ssw3 dictyosteli
128	17	1.3	1942	5	Q8MYF9	Q8myf9 dictyosteli
129	17	1.3	1969	5	O15763	O15763 dictyosteli
130	17	1.3	2015	5	Q9U5Y1	Q9u5y1 dictyosteli
131	17	1.3	2016	5	Q8ILY1	Q8ily1 plasmodium
132	17	1.3	2084	5	Q8ILJ6	Q8ilj6 plasmodium
133	17	1.3	2184	5	Q8ILQ6	Q8ilq6 plasmodium
134	17	1.3	2230	5	Q8MNM0	Q8mnm0 dictyosteli
135	17	1.3	2235	5	Q8IB09	Q8ib09 plasmodium
136	17	1.3	2273	5	Q8MN83	Q8mn83 dictyosteli
137	17	1.3	2344	5	Q8SSV9	Q8ssv9 dictyosteli
138	17	1.3	2359	5	Q8I297	Q8i297 plasmodium
139	17	1.3	2388	5	Q8MYG5	Q8myg5 dictyosteli
140	17	1.3	2569	5	Q8IBG8	Q8ibg8 plasmodium
141	17	1.3	2577	5	Q8IAM1	Q8iam1 plasmodium
142	17	1.3	2625	5	Q8MMZ9	Q8mmz9 dictyosteli
143	17	1.3	2722	5	Q8MP57	Q8mp57 dictyosteli
144	17	1.3	2752	5	Q8ICT8	Q8ict8 plasmodium
145	17	1.3	2763	5	Q8II48	Q8ii48 plasmodium
146	17	1.3	2763	5	Q8I3X5	Q8i3x5 plasmodium
147	17	1.3	2792	5	Q8I552	Q8i552 plasmodium
148	17	1.3	2874	5	Q8ICX0	Q8icx0 plasmodium
149	17	1.3	2964	5	Q8I7W7	Q8i7w7 dictyosteli
150	17	1.3	3178	5	Q8I2P8	Q8i2p8 plasmodium
151	17	1.3	3193	5	Q8I590	Q8i590 plasmodium
152	17	1.3	3207	5	Q8IHQ0	Q8ihq0 plasmodium
153	17	1.3	3265	5	Q8ID63	Q8id63 plasmodium
154	17	1.3	3268	5	Q8IC30	Q8ic30 plasmodium
155	17	1.3	3334	5	Q8IIS9	Q8iis9 plasmodium
156	17	1.3	3371	5	Q8IIW4	Q8iiw4 plasmodium
157	17	1.3	3417	5	Q8T1I6	Q8t1i6 dictyosteli
158	17	1.3	3432	5	Q95PH9	Q95ph9 dictyosteli
159	17	1.3	3933	5	O97239	O97239 plasmodium
160	17	1.3	4138	5	Q8I1Y3	Q8i1y3 plasmodium
161	17	1.3	4202	5	Q8I2S8	Q8i2s8 plasmodium
162	17	1.3	4466	5	Q8ILS4	Q8ils4 plasmodium
163	17	1.3	4550	5	O77336	O77336 plasmodium
164	17	1.3	4780	5	Q8IE54	Q8ie54 plasmodium
165	17	1.3	5561	5	Q8IBH6	Q8ibh6 plasmodium
166	17	1.3	8591	5	Q8IB94	Q8ib94 plasmodium
167	17	1.3	10061	5	Q8I3Z1	Q8i3z1 plasmodium
168	16	1.2	65	5	Q8MXQ3	Q8mxq3 dictyosteli
169	16	1.2	68	5	Q8MN60	Q8mn60 dictyosteli
170	16	1.2	79	5	Q8MNB2	Q8mnb2 dictyosteli
171	16	1.2	86	5	Q8MN69	Q8mn69 dictyosteli
172	16	1.2	88	5	Q8MN24	Q8mn24 dictyosteli
173	16	1.2	104	5	Q8T225	Q8t225 dictyosteli
174	16	1.2	130	10	Q8S2B4	Q8s2b4 oryza sativ
175	16	1.2	135	17	Q8PS74	Q8ps74 methanosarc
176	16	1.2	150	10	Q9FH22	Q9fh22 arabidopsis
177	16	1.2	158	5	Q8MNP2	Q8mnp2 dictyosteli
178	16	1.2	160	11	O35058	O35058 mus musculu
179	16	1.2	166	5	Q8MNQ0	Q8mnq0 dictyosteli
180	16	1.2	185	5	Q8MN57	Q8mn57 dictyosteli
181	16	1.2	187	5	Q8T279	Q8t279 dictyosteli
182	16	1.2	188	10	O64868	O64868 arabidopsis
183	16	1.2	198	5	Q8MN88	Q8mn88 dictyosteli
184	16	1.2	212	5	Q8T1D8	Q8t1d8 dictyosteli
185	16	1.2	229	5	Q8MP31	Q8mp31 dictyosteli
186	16	1.2	238	5	Q8T2C1	Q8t2c1 dictyosteli
187	16	1.2	252	5	Q8T227	Q8t227 dictyosteli
188	16	1.2	259	5	O96550	O96550 dictyosteli
189	16	1.2	261	5	Q8T868	Q8t868 dictyosteli
190	16	1.2	268	5	Q8T862	Q8t862 dictyosteli
191	16	1.2	275	5	Q8SSQ4	Q8ssq4 dictyosteli
192	16	1.2	282	5	Q8MXL8	Q8mxl8 dictyosteli
193	16	1.2	301	5	Q8T1N1	Q8t1n1 dictyosteli
194	16	1.2	302	5	Q8ILP7	Q8ilp7 plasmodium
195	16	1.2	305	5	Q8MN18	Q8mn18 dictyosteli
196	16	1.2	308	5	Q25778	Q25778 plasmodium
197	16	1.2	312	5	Q8MNJ8	Q8mnj8 dictyosteli
198	16	1.2	312	5	Q8MNA4	Q8mna4 dictyosteli
199	16	1.2	316	5	Q8MNF2	Q8mnf2 dictyosteli
200	16	1.2	317	5	O97336	O97336 plasmodium
201	16	1.2	322	5	Q8T1V7	Q8t1v7 dictyosteli
202	16	1.2	323	5	Q8ICG7	Q8icg7 plasmodium
203	16	1.2	328	10	O24463	O24463 zea mays (m
204	16	1.2	340	5	Q8T212	Q8t212 dictyosteli
205	16	1.2	341	5	Q8T1C1	Q8t1c1 dictyosteli
206	16	1.2	349	5	Q8MN85	Q8mn85 dictyosteli
207	16	1.2	351	5	O96548	O96548 dictyosteli
208	16	1.2	357	5	Q8I2S1	Q8i2s1 plasmodium
209	16	1.2	369	5	Q8T1D9	Q8t1d9 dictyosteli
210	16	1.2	372	5	Q8IJV8	Q8ijv8 plasmodium
211	16	1.2	373	8	Q35905	Q35905 saccharomyc
212	16	1.2	373	10	Q9SMX7	Q9smx7 arabidopsis
213	16	1.2	374	5	Q8T1K2	Q8t1k2 dictyosteli
214	16	1.2	383	5	Q8MNL9	Q8mnl9 dictyosteli
215	16	1.2	398	3	Q9Y764	Q9y764 candida alb
216	16	1.2	399	5	Q8MNB4	Q8mnb4 dictyosteli
217	16	1.2	401	5	Q27229	Q27229 dictyosteli
218	16	1.2	405	5	Q8MNA0	Q8mna0 dictyosteli
219	16	1.2	406	5	Q8MMQ0	Q8mmq0 dictyosteli
220	16	1.2	407	5	Q8T831	Q8t831 dictyosteli
221	16	1.2	414	5	Q8T2J5	Q8t2j5 dictyosteli
222	16	1.2	417	5	Q8MP36	Q8mp36 dictyosteli
223	16	1.2	417	17	Q8PTG9	Q8ptg9 methanosarc
224	16	1.2	421	5	Q8MTX3	Q8mtx3 theromyzon
225	16	1.2	423	5	Q8MMP6	Q8mmp6 dictyosteli
226	16	1.2	423	5	Q967M5	Q967m5 dictyosteli
227	16	1.2	425	5	O15755	O15755 dictyosteli
228	16	1.2	425	5	Q8T1K4	Q8t1k4 dictyosteli
229	16	1.2	426	5	Q8T828	Q8t828 dictyosteli
230	16	1.2	428	3	Q12646	Q12646 neocallimas
231	16	1.2	428	5	Q8T2C5	Q8t2c5 dictyosteli
232	16	1.2	432	5	Q8IRS8	Q8irs8 drosophila
233	16	1.2	437	5	Q8SSX8	Q8ssx8 dictyosteli
234	16	1.2	437	5	Q8T199	Q8t199 dictyosteli
235	16	1.2	442	5	Q8SSQ9	Q8ssq9 dictyosteli

236	16	1.2	450	5	Q8MP78	Q8mp78 dictyosteli
237	16	1.2	457	5	Q8T283	Q8t283 dictyosteli
238	16	1.2	459	5	Q8I562	Q8i562 plasmodium
239	16	1.2	463	5	Q8MMR7	Q8mmr7 dictyosteli
240	16	1.2	463	5	Q8IIR4	Q8iir4 plasmodium
241	16	1.2	474	5	Q8SR35	Q8sr35 encephalito
242	16	1.2	476	5	Q8T1Y3	Q8t1y3 dictyosteli
243	16	1.2	490	5	Q8T2J2	Q8t2j2 dictyosteli
244	16	1.2	490	5	Q8T2R4	Q8t2r4 dictyosteli
245	16	1.2	498	5	Q8T1H4	Q8t1h4 dictyosteli
246	16	1.2	508	5	Q8IDS4	Q8ids4 plasmodium
247	16	1.2	513	5	Q8T664	Q8t664 dictyosteli
248	16	1.2	520	5	Q8T1J4	Q8t1j4 dictyosteli
249	16	1.2	520	5	Q8I5N1	Q8i5n1 plasmodium
250	16	1.2	522	5	Q8I6Y6	Q8i6y6 dictyosteli
251	16	1.2	527	5	Q8T1V9	Q8t1v9 dictyosteli
252	16	1.2	530	5	Q8T2A0	Q8t2a0 dictyosteli
253	16	1.2	534	5	O43989	O43989 dictyosteli
254	16	1.2	535	5	Q8T2U6	Q8t2u6 dictyosteli
255	16	1.2	538	5	Q8SSY9	Q8ssy9 dictyosteli
256	16	1.2	540	5	O76771	O76771 dictyosteli
257	16	1.2	540	5	Q8T1Y8	Q8t1y8 dictyosteli
258	16	1.2	542	5	Q8T1R3	Q8t1r3 dictyosteli
259	16	1.2	547	5	Q8ID99	Q8id99 plasmodium
260	16	1.2	548	5	Q8T1H3	Q8t1h3 dictyosteli
261	16	1.2	548	5	Q8T1S2	Q8t1s2 dictyosteli
262	16	1.2	559	5	Q9GPT1	Q9gpt1 dictyosteli
263	16	1.2	559	5	Q8T1R1	Q8t1r1 dictyosteli
264	16	1.2	565	5	O15754	O15754 dictyosteli
265	16	1.2	566	5	Q23910	Q23910 dictyosteli
266	16	1.2	566	5	Q8T285	Q8t285 dictyosteli
267	16	1.2	566	5	Q8MND0	Q8mnd0 dictyosteli
268	16	1.2	569	2	Q9AJB2	Q9ajb2 ruminococcu
269	16	1.2	571	5	Q8T2A7	Q8t2a7 dictyosteli
270	16	1.2	572	5	Q8MNK9	Q8mnk9 dictyosteli
271	16	1.2	575	5	O77334	O77334 plasmodium
272	16	1.2	577	5	Q8MN28	Q8mn28 dictyosteli
273	16	1.2	578	5	Q8ICM3	Q8icm3 plasmodium
274	16	1.2	579	5	O60950	O60950 dictyosteli
275	16	1.2	585	5	Q8ILQ8	Q8ilq8 plasmodium
276	16	1.2	586	5	Q8T2J6	Q8t2j6 dictyosteli
277	16	1.2	595	5	Q8T2Q4	Q8t2q4 dictyosteli
278	16	1.2	605	5	Q8SSZ9	Q8ssz9 dictyosteli
279	16	1.2	609	5	Q9NGW8	Q9ngw8 dictyosteli
280	16	1.2	616	5	Q8T1D3	Q8t1d3 dictyosteli
281	16	1.2	619	5	Q8MNB6	Q8mnb6 dictyosteli
282	16	1.2	620	5	Q8T2C0	Q8t2c0 dictyosteli
283	16	1.2	628	5	Q9VSB3	Q9vsb3 drosophila
284	16	1.2	628	5	Q8ILB8	Q8ilb8 plasmodium
285	16	1.2	629	5	Q8T2A9	Q8t2a9 dictyosteli
286	16	1.2	630	5	Q8IB88	Q8ib88 plasmodium
287	16	1.2	637	5	Q8MND3	Q8mnd3 dictyosteli
288	16	1.2	637	5	Q8IB92	Q8ib92 plasmodium
289	16	1.2	637	5	Q8I1Z4	Q8i1z4 plasmodium
290	16	1.2	649	5	Q9VW88	Q9vw88 drosophila
291	16	1.2	655	5	Q8I4R6	Q8i4r6 plasmodium
292	16	1.2	656	5	Q8ICP6	Q8icp6 plasmodium
293	16	1.2	659	5	Q8T1R0	Q8t1r0 dictyosteli
294	16	1.2	660	5	Q94485	Q94485 dictyosteli
295	16	1.2	660	5	O00885	O00885 dictyosteli
296	16	1.2	665	5	Q8T208	Q8t208 dictyosteli
297	16	1.2	665	5	Q8I3I8	Q8i3i8 plasmodium
298	16	1.2	666	5	Q8IM66	Q8im66 plasmodium
299	16	1.2	667	5	Q9U0K7	Q9u0k7 plasmodium
300	16	1.2	680	5	Q8IJI8	Q8iji8 plasmodium
301	16	1.2	684	5	Q8I3M7	Q8i3m7 plasmodium
302	16	1.2	685	5	Q8ID39	Q8id39 plasmodium
303	16	1.2	686	5	Q8MNJ2	Q8mnj2 dictyosteli
304	16	1.2	686	5	O96245	O96245 plasmodium
305	16	1.2	687	5	Q8IE27	Q8ie27 plasmodium
306	16	1.2	692	5	Q8MN01	Q8mn01 dictyosteli
307	16	1.2	693	5	Q8MXN4	Q8mxn4 dictyosteli
308	16	1.2	694	5	Q8T2F2	Q8t2f2 dictyosteli
309	16	1.2	700	5	Q8T6I6	Q8t6i6 dictyosteli
310	16	1.2	705	5	Q8T182	Q8t182 dictyosteli
311	16	1.2	714	5	Q8MN30	Q8mn30 dictyosteli
312	16	1.2	714	5	Q8IIU9	Q8iiu9 plasmodium
313	16	1.2	715	5	Q8IKY0	Q8iky0 plasmodium
314	16	1.2	718	5	Q9XYP6	Q9xyp6 dictyosteli
315	16	1.2	720	3	Q9P3M3	Q9p3m3 neurospora
316	16	1.2	721	5	Q8T218	Q8t218 dictyosteli
317	16	1.2	722	5	Q8T232	Q8t232 dictyosteli
318	16	1.2	726	2	Q9AJB3	Q9ajb3 ruminococcu
319	16	1.2	733	5	Q8T2R6	Q8t2r6 dictyosteli
320	16	1.2	743	5	Q9NGX0	Q9ngx0 dictyosteli
321	16	1.2	747	5	Q8MMQ3	Q8mmq3 dictyosteli
322	16	1.2	757	5	Q9GSG6	Q9gsg6 dictyosteli
323	16	1.2	768	5	Q9NAX6	Q9nax6 dictyosteli
324	16	1.2	772	5	Q8MNI4	Q8mni4 dictyosteli
325	16	1.2	779	5	Q8IBN6	Q8ibn6 plasmodium
326	16	1.2	782	5	Q8MMV7	Q8mmv7 dictyosteli
327	16	1.2	786	5	Q8IDS3	Q8ids3 plasmodium
328	16	1.2	791	5	O00841	O00841 dictyosteli
329	16	1.2	798	5	Q8T200	Q8t200 dictyosteli
330	16	1.2	800	5	O43988	O43988 dictyosteli
331	16	1.2	802	5	Q8I1R5	Q8i1r5 plasmodium
332	16	1.2	803	5	Q8T2J0	Q8t2j0 dictyosteli
333	16	1.2	803	5	Q8IEC5	Q8iec5 plasmodium
334	16	1.2	807	5	Q8IB74	Q8ib74 plasmodium
335	16	1.2	812	5	Q8MN39	Q8mn39 dictyosteli
336	16	1.2	818	5	Q8I5D6	Q8i5d6 plasmodium
337	16	1.2	822	5	Q8I5B3	Q8i5b3 plasmodium
338	16	1.2	823	5	Q8MNG7	Q8mng7 dictyosteli
339	16	1.2	824	5	O00911	O00911 dictyosteli
340	16	1.2	827	5	Q8IEM2	Q8iem2 plasmodium
341	16	1.2	831	5	O15756	O15756 dictyosteli
342	16	1.2	832	5	Q8T2K2	Q8t2k2 dictyosteli
343	16	1.2	837	5	Q8MYE9	Q8mye9 dictyosteli
344	16	1.2	851	5	Q94488	Q94488 dictyosteli
345	16	1.2	852	5	Q8SSS7	Q8sss7 dictyosteli
346	16	1.2	853	5	Q8T4K1	Q8t4k1 dictyosteli
347	16	1.2	855	5	Q8SSV6	Q8ssv6 dictyosteli
348	16	1.2	860	5	Q8IK95	Q8ik95 plasmodium
349	16	1.2	865	5	Q8IKP3	Q8ikp3 plasmodium
350	16	1.2	872	5	P90523	P90523 dictyosteli
351	16	1.2	876	5	Q8SSS4	Q8sss4 dictyosteli
352	16	1.2	877	5	Q23853	Q23853 dictyosteli
353	16	1.2	882	5	Q8II45	Q8ii45 plasmodium
354	16	1.2	885	5	Q8SSZ7	Q8ssz7 dictyosteli
355	16	1.2	893	5	Q8T7Y3	Q8t7y3 dictyosteli
356	16	1.2	895	5	Q8I8U2	Q8i8u2 dictyosteli
357	16	1.2	898	5	Q8T1G7	Q8t1g7 dictyosteli
358	16	1.2	901	5	Q8T806	Q8t806 dictyosteli
359	16	1.2	911	5	Q8MXP0	Q8mxp0 dictyosteli
360	16	1.2	919	5	Q8IJ92	Q8ij92 plasmodium
361	16	1.2	934	5	Q8SSV2	Q8ssv2 dictyosteli
362	16	1.2	937	5	Q8MND6	Q8mnd6 dictyosteli
363	16	1.2	947	5	O15739	O15739 dictyosteli
364	16	1.2	952	5	Q8MMM4	Q8mmm4 dictyosteli
365	16	1.2	955	5	Q8IDT3	Q8idt3 plasmodium
366	16	1.2	964	5	Q8T856	Q8t856 dictyosteli
367	16	1.2	969	5	Q9XZ11	Q9xz11 drosophila
368	16	1.2	970	5	Q8T143	Q8t143 dictyosteli
369	16	1.2	986	5	Q8MMQ8	Q8mmq8 dictyosteli
370	16	1.2	990	5	Q8I437	Q8i437 plasmodium
371	16	1.2	992	5	Q8I1Q4	Q8i1q4 plasmodium
372	16	1.2	995	5	Q8IJ91	Q8ij91 plasmodium
373	16	1.2	1004	5	Q8MP26	Q8mp26 dictyosteli
374	16	1.2	1017	5	Q8MN02	Q8mn02 dictyosteli
375	16	1.2	1022	5	Q8MMY9	Q8mmy9 dictyosteli
376	16	1.2	1022	5	Q8IET1	Q8iet1 plasmodium
377	16	1.2	1024	5	Q8MMW3	Q8mmw3 dictyosteli
378	16	1.2	1032	5	Q8I524	Q8i524 plasmodium
379	16	1.2	1033	5	Q8MNN0	Q8mnn0 dictyosteli
380	16	1.2	1036	5	Q8I355	Q8i355 plasmodium
381	16	1.2	1037	5	Q8IS18	Q8is18 dictyosteli

382	16	1.2	1043	5	Q8MNB7	Q8mnb7 dictyosteli
383	16	1.2	1050	5	Q8IIQ6	Q8iiq6 plasmodium
384	16	1.2	1064	5	Q25773	Q25773 plasmodium
385	16	1.2	1075	3	Q12221	Q12221 saccharomyc
386	16	1.2	1080	5	Q8T1T2	Q8t1t2 dictyosteli
387	16	1.2	1090	5	Q8MMY2	Q8mmy2 dictyosteli
388	16	1.2	1099	5	Q9U9K8	Q9u9k8 dictyosteli
389	16	1.2	1104	5	Q8MMM1	Q8mmm1 dictyosteli
390	16	1.2	1104	5	Q8IJG4	Q8ijg4 plasmodium
391	16	1.2	1108	5	Q8T2E2	Q8t2e2 dictyosteli
392	16	1.2	1112	5	Q8MMS7	Q8mms7 dictyosteli
393	16	1.2	1118	5	Q8T1I4	Q8t1i4 dictyosteli
394	16	1.2	1123	5	Q8I811	Q8i811 dictyosteli
395	16	1.2	1126	5	Q8T278	Q8t278 dictyosteli
396	16	1.2	1140	5	Q8T1T0	Q8t1t0 dictyosteli
397	16	1.2	1145	5	Q8IIS6	Q8iis6 plasmodium
398	16	1.2	1147	5	Q95PH8	Q95ph8 dictyosteli
399	16	1.2	1149	5	Q8ILP5	Q8ilp5 plasmodium
400	16	1.2	1152	5	Q8T216	Q8t216 dictyosteli
401	16	1.2	1156	5	Q8I474	Q8i474 plasmodium
402	16	1.2	1157	5	O00895	O00895 plasmodium
403	16	1.2	1165	5	Q8IB76	Q8ib76 plasmodium
404	16	1.2	1166	5	Q8ILR0	Q8ilr0 plasmodium
405	16	1.2	1172	5	Q8IB93	Q8ib93 plasmodium
406	16	1.2	1176	5	Q8SSV4	Q8ssv4 dictyosteli
407	16	1.2	1185	5	Q8ICL1	Q8icl1 plasmodium
408	16	1.2	1190	5	Q8IJZ6	Q8ijz6 plasmodium
409	16	1.2	1191	5	Q9XYY3	Q9xyy3 dictyosteli
410	16	1.2	1192	5	Q8IEI5	Q8iei5 plasmodium
411	16	1.2	1196	5	Q23866	Q23866 dictyosteli
412	16	1.2	1197	5	Q8I391	Q8i391 plasmodium
413	16	1.2	1200	5	O97277	O97277 plasmodium
414	16	1.2	1203	5	Q8T863	Q8t863 dictyosteli
415	16	1.2	1204	5	Q8T134	Q8t134 dictyosteli
416	16	1.2	1212	5	Q9U0L0	Q9u0l0 plasmodium
417	16	1.2	1213	5	Q95PH2	Q95ph2 dictyosteli
418	16	1.2	1221	5	Q8IBN0	Q8ibn0 plasmodium
419	16	1.2	1223	5	Q8ST14	Q8st14 dictyosteli
420	16	1.2	1225	5	O15784	O15784 dictyosteli
421	16	1.2	1225	5	Q95PI2	Q95pi2 dictyosteli
422	16	1.2	1236	5	Q8ICJ8	Q8icj8 plasmodium
423	16	1.2	1245	5	Q8I2R3	Q8i2r3 plasmodium
424	16	1.2	1250	5	Q8I611	Q8i611 plasmodium
425	16	1.2	1252	5	Q8IIG0	Q8iig0 plasmodium
426	16	1.2	1256	5	Q25770	Q25770 plasmodium
427	16	1.2	1262	5	Q8MN40	Q8mn40 dictyosteli
428	16	1.2	1271	5	O15749	O15749 dictyosteli
429	16	1.2	1287	5	Q9NGS5	Q9ngs5 dictyosteli
430	16	1.2	1289	5	Q8IBA2	Q8iba2 plasmodium
431	16	1.2	1290	5	Q9U6L1	Q9u6l1 plasmodium
432	16	1.2	1297	5	Q8T2U8	Q8t2u8 dictyosteli
433	16	1.2	1298	5	Q9U0I0	Q9u0i0 plasmodium
434	16	1.2	1301	5	Q8WSK5	Q8wsk5 plasmodium
435	16	1.2	1301	5	Q8IHQ2	Q8ihq2 plasmodium
436	16	1.2	1304	5	Q8T126	Q8t126 dictyosteli
437	16	1.2	1304	5	Q8IKW2	Q8ikw2 plasmodium
438	16	1.2	1312	5	Q8MVR0	Q8mvr0 dictyosteli
439	16	1.2	1313	5	Q8T1M7	Q8t1m7 dictyosteli
440	16	1.2	1318	5	Q95PH4	Q95ph4 dictyosteli
441	16	1.2	1318	5	Q8MMX3	Q8mmx3 dictyosteli
442	16	1.2	1320	5	Q8ILJ3	Q8ilj3 plasmodium
443	16	1.2	1324	5	Q8T2N1	Q8t2n1 dictyosteli
444	16	1.2	1324	5	Q8IFQ8	Q8ifq8 plasmodium
445	16	1.2	1326	5	Q8IB98	Q8ib98 plasmodium
446	16	1.2	1331	5	Q8IBF6	Q8ibf6 plasmodium
447	16	1.2	1336	5	Q8IIC8	Q8iic8 plasmodium
448	16	1.2	1337	5	Q8IC23	Q8ic23 plasmodium
449	16	1.2	1338	5	Q8MN23	Q8mn23 dictyosteli
450	16	1.2	1340	5	Q8ICF1	Q8icf1 plasmodium
451	16	1.2	1342	5	Q8I2I8	Q8i2i8 plasmodium
452	16	1.2	1348	5	Q8T234	Q8t234 dictyosteli
453	16	1.2	1359	5	Q8T6G6	Q8t6g6 dictyosteli
454	16	1.2	1364	5	Q8T223	Q8t223 dictyosteli

455	16	1.2	1364	5	Q8MN00	Q8mn00 dictyosteli
456	16	1.2	1368	5	Q9V6J0	Q9v6j0 drosophila
457	16	1.2	1373	5	Q8I3A8	Q8i3a8 plasmodium
458	16	1.2	1374	5	Q8IKR9	Q8ikr9 plasmodium
459	16	1.2	1377	5	Q8I5S6	Q8i5s6 plasmodium
460	16	1.2	1380	5	Q8MN90	Q8mn90 dictyosteli
461	16	1.2	1390	5	Q8T6H6	Q8t6h6 dictyosteli
462	16	1.2	1390	5	O77033	O77033 dictyosteli
463	16	1.2	1393	5	Q8T690	Q8t690 dictyosteli
464	16	1.2	1394	5	Q8ILG6	Q8ilg6 plasmodium
465	16	1.2	1394	5	Q8IDL6	Q8idl6 plasmodium
466	16	1.2	1402	5	O96668	O96668 dictyosteli
467	16	1.2	1404	5	Q8MP67	Q8mp67 dictyosteli
468	16	1.2	1405	5	Q8MNC0	Q8mnc0 dictyosteli
469	16	1.2	1419	5	Q8I3Z3	Q8i3z3 plasmodium
470	16	1.2	1431	5	Q8MN11	Q8mn11 dictyosteli
471	16	1.2	1432	5	Q8T9W4	Q8t9w4 dictyosteli
472	16	1.2	1434	5	Q8IJI3	Q8iji3 plasmodium
473	16	1.2	1434	5	Q8IJ77	Q8ij77 plasmodium
474	16	1.2	1457	5	Q8IS20	Q8is20 dictyosteli
475	16	1.2	1461	5	Q8ST04	Q8st04 dictyosteli
476	16	1.2	1464	5	Q8IIF6	Q8iif6 plasmodium
477	16	1.2	1464	5	Q8IHS3	Q8ihs3 plasmodium
478	16	1.2	1472	5	Q8T1D7	Q8t1d7 dictyosteli
479	16	1.2	1476	5	Q8T676	Q8t676 dictyosteli
480	16	1.2	1476	5	Q8ST66	Q8st66 dictyosteli
481	16	1.2	1476	5	Q965D3	Q965d3 dictyosteli
482	16	1.2	1478	5	Q8IKQ3	Q8ikq3 plasmodium
483	16	1.2	1484	5	Q8IB67	Q8ib67 plasmodium
484	16	1.2	1486	5	Q9XZS0	Q9xzs0 dictyosteli
485	16	1.2	1488	5	Q8IDP7	Q8idp7 plasmodium
486	16	1.2	1492	5	Q8SSU1	Q8ssu1 dictyosteli
487	16	1.2	1509	5	Q8T688	Q8t688 dictyosteli
488	16	1.2	1520	5	Q8T687	Q8t687 dictyosteli
489	16	1.2	1529	5	Q9GQC2	Q9gqc2 dictyosteli
490	16	1.2	1529	5	Q8IS21	Q8is21 dictyosteli
491	16	1.2	1551	5	Q8IS13	Q8is13 dictyosteli
492	16	1.2	1552	5	Q8I3N4	Q8i3n4 plasmodium
493	16	1.2	1556	5	Q8ST25	Q8st25 dictyosteli
494	16	1.2	1558	5	Q8ILL5	Q8ill5 plasmodium
495	16	1.2	1559	5	Q8IDC3	Q8idc3 plasmodium
496	16	1.2	1560	5	Q9GRX5	Q9grx5 dictyosteli
497	16	1.2	1561	5	Q9U987	Q9u987 dictyosteli
498	16	1.2	1587	5	Q8IC25	Q8ic25 plasmodium
499	16	1.2	1593	5	Q8T1T7	Q8t1t7 dictyosteli
500	16	1.2	1597	5	Q8IJW6	Q8ijw6 plasmodium
501	16	1.2	1604	5	Q8MMW7	Q8mmw7 dictyosteli
502	16	1.2	1606	5	Q8IDH6	Q8idh6 plasmodium
503	16	1.2	1612	5	Q8IFN0	Q8ifn0 plasmodium
504	16	1.2	1618	5	Q8I5I3	Q8i5i3 plasmodium
505	16	1.2	1619	5	O77382	O77382 plasmodium
506	16	1.2	1632	5	Q8IKM7	Q8ikm7 plasmodium
507	16	1.2	1641	5	Q8IKZ0	Q8ikz0 plasmodium
508	16	1.2	1661	5	Q8IE71	Q8ie71 plasmodium
509	16	1.2	1662	5	Q8I569	Q8i569 plasmodium
510	16	1.2	1681	5	Q8IJ39	Q8ij39 plasmodium
511	16	1.2	1697	5	Q8IJA6	Q8ija6 plasmodium
512	16	1.2	1699	5	Q95PH3	Q95ph3 dictyosteli
513	16	1.2	1702	5	Q9GTN7	Q9gtn7 dictyosteli
514	16	1.2	1702	5	Q8IM14	Q8im14 plasmodium
515	16	1.2	1728	5	Q8I307	Q8i307 plasmodium
516	16	1.2	1754	5	Q8I2Q1	Q8i2q1 plasmodium
517	16	1.2	1770	5	Q8I3A3	Q8i3a3 plasmodium
518	16	1.2	1774	5	Q8IAU8	Q8iau8 plasmodium
519	16	1.2	1792	5	Q8ID94	Q8id94 plasmodium
520	16	1.2	1811	5	Q8IJD3	Q8ijd3 plasmodium
521	16	1.2	1817	5	O96253	O96253 plasmodium
522	16	1.2	1826	5	Q8I2Y3	Q8i2y3 plasmodium
523	16	1.2	1828	5	Q8IIS4	Q8iis4 plasmodium
524	16	1.2	1831	5	Q8I5T8	Q8i5t8 plasmodium
525	16	1.2	1833	5	Q8IAZ5	Q8iaz5 plasmodium
526	16	1.2	1838	5	Q8T1N5	Q8t1n5 dictyosteli
527	16	1.2	1844	5	O96201	O96201 plasmodium

528	16	1.2	1855	5	Q9TX75	Q9tx75 plasmodium
529	16	1.2	1855	5	Q9BHN0	Q9bhn0 plasmodium
530	16	1.2	1868	5	Q8IAL9	Q8ial9 plasmodium
531	16	1.2	1898	5	Q8ILA3	Q8ila3 plasmodium
532	16	1.2	1912	5	Q9U0H1	Q9u0h1 plasmodium
533	16	1.2	1916	5	Q8IBC1	Q8ibc1 plasmodium
534	16	1.2	1922	5	Q8I2P4	Q8i2p4 plasmodium
535	16	1.2	1923	5	Q8IBH8	Q8ibh8 plasmodium
536	16	1.2	1929	5	Q8IHL2	Q8ihl2 dictyosteli
537	16	1.2	1936	5	Q8I5A6	Q8i5a6 plasmodium
538	16	1.2	1951	5	Q8ILY5	Q8ily5 plasmodium
539	16	1.2	1988	5	Q8IIT8	Q8iit8 plasmodium
540	16	1.2	1989	5	Q8I4U7	Q8i4u7 plasmodium
541	16	1.2	1997	5	Q8I5L6	Q8i5l6 plasmodium
542	16	1.2	2010	5	Q8ICX5	Q8icx5 plasmodium
543	16	1.2	2062	5	Q95PH6	Q95ph6 dictyosteli
544	16	1.2	2062	5	Q8IBP4	Q8ibp4 plasmodium
545	16	1.2	2123	5	Q9U9S7	Q9u9s7 dictyosteli
546	16	1.2	2130	5	Q8IBC6	Q8ibc6 plasmodium
547	16	1.2	2147	5	Q8IIH3	Q8iih3 plasmodium
548	16	1.2	2148	5	Q8I1N6	Q8i1n6 plasmodium
549	16	1.2	2150	5	Q23863	Q23863 dictyosteli
550	16	1.2	2159	5	Q8ST00	Q8st00 dictyosteli
551	16	1.2	2190	5	Q8IBL5	Q8ibl5 plasmodium
552	16	1.2	2206	5	Q8IBU8	Q8ibu8 plasmodium
553	16	1.2	2221	5	Q8I259	Q8i259 plasmodium
554	16	1.2	2254	5	Q8T1Y5	Q8t1y5 dictyosteli
555	16	1.2	2275	5	Q8IHV8	Q8ihv8 plasmodium
556	16	1.2	2283	5	Q8IC35	Q8ic35 plasmodium
557	16	1.2	2284	5	Q8I5Y7	Q8i5y7 plasmodium
558	16	1.2	2353	5	Q8ILH9	Q8ilh9 plasmodium
559	16	1.2	2379	5	Q8IJW3	Q8ijw3 plasmodium
560	16	1.2	2380	5	O96266	O96266 plasmodium
561	16	1.2	2399	5	Q8IBB0	Q8ibb0 plasmodium
562	16	1.2	2404	5	Q8ICD5	Q8icd5 plasmodium
563	16	1.2	2405	5	Q8IJL2	Q8ijl2 plasmodium
564	16	1.2	2415	5	Q8I1X1	Q8i1x1 plasmodium
565	16	1.2	2432	5	Q8I5I9	Q8i5i9 plasmodium
566	16	1.2	2454	5	Q8T2G3	Q8t2g3 dictyosteli
567	16	1.2	2467	5	Q8IID3	Q8iid3 plasmodium
568	16	1.2	2472	5	Q8T2M5	Q8t2m5 dictyosteli
569	16	1.2	2472	5	Q8I1P3	Q8i1p3 plasmodium
570	16	1.2	2515	5	O77365	O77365 plasmodium
571	16	1.2	2518	5	Q8IEH2	Q8ieh2 plasmodium
572	16	1.2	2535	5	Q8I403	Q8i403 plasmodium
573	16	1.2	2578	5	Q8IJP9	Q8ijp9 plasmodium
574	16	1.2	2612	5	Q8I5X5	Q8i5x5 plasmodium
575	16	1.2	2668	5	Q8IE42	Q8ie42 plasmodium
576	16	1.2	2672	5	Q8IEE4	Q8iee4 plasmodium
577	16	1.2	2700	5	Q8ILC5	Q8ilc5 plasmodium
578	16	1.2	2756	10	Q9LJ60	Q9lj60 arabidopsis
579	16	1.2	2770	5	Q8ILV0	Q8ilv0 plasmodium
580	16	1.2	2773	5	Q8IBF4	Q8ibf4 plasmodium
581	16	1.2	2792	5	Q8I4R2	Q8i4r2 plasmodium
582	16	1.2	2814	5	Q8IC15	Q8ic15 plasmodium
583	16	1.2	2843	5	Q963L5	Q963l5 dictyosteli
584	16	1.2	2849	5	Q8IHY4	Q8ihy4 plasmodium
585	16	1.2	2907	5	Q8IC71	Q8ic71 plasmodium
586	16	1.2	2910	5	Q8IBY8	Q8iby8 plasmodium
587	16	1.2	2932	5	Q8I4T6	Q8i4t6 plasmodium
588	16	1.2	2940	5	Q8IHP9	Q8ihp9 plasmodium
589	16	1.2	2943	5	Q8IK94	Q8ik94 plasmodium
590	16	1.2	3099	5	Q8MYH0	Q8myh0 dictyosteli
591	16	1.2	3134	5	Q8I414	Q8i414 plasmodium
592	16	1.2	3135	5	Q8I3B5	Q8i3b5 plasmodium
593	16	1.2	3209	5	Q8I5D0	Q8i5d0 plasmodium
594	16	1.2	3218	5	Q8ILX0	Q8ilx0 plasmodium
595	16	1.2	3223	5	Q8IKJ2	Q8ikj2 plasmodium
596	16	1.2	3248	5	Q8ICP9	Q8icp9 plasmodium
597	16	1.2	3303	5	Q8I339	Q8i339 plasmodium
598	16	1.2	3364	5	Q8IM60	Q8im60 plasmodium
599	16	1.2	3468	5	Q8II04	Q8ii04 plasmodium
600	16	1.2	3482	5	Q8ID46	Q8id46 plasmodium
601	16	1.2	3569	5	Q8IBV8	Q8ibv8 plasmodium
602	16	1.2	3597	5	Q8ILR5	Q8ilr5 plasmodium
603	16	1.2	3620	5	Q968T6	Q968t6 plasmodium
604	16	1.2	3633	5	Q8IHL0	Q8ihl0 dictyosteli
605	16	1.2	3763	5	Q8T2A1	Q8t2a1 dictyosteli
606	16	1.2	3787	5	Q8I1Z6	Q8i1z6 plasmodium
607	16	1.2	3844	5	Q94648	Q94648 plasmodium
608	16	1.2	3848	5	O76737	O76737 dictyosteli
609	16	1.2	3848	5	Q8IDR0	Q8idr0 plasmodium
610	16	1.2	4095	5	Q8ICN0	Q8icn0 plasmodium
611	16	1.2	4226	5	Q9N9H5	Q9n9h5 plasmodium
612	16	1.2	4226	5	Q8IHY1	Q8ihy1 plasmodium
613	16	1.2	4405	5	Q8ILZ2	Q8ilz2 plasmodium
614	16	1.2	4431	5	Q8IJI6	Q8iji6 plasmodium
615	16	1.2	4493	5	Q8MPA9	Q8mpa9 dictyosteli
616	16	1.2	4494	5	Q8I5I2	Q8i5i2 plasmodium
617	16	1.2	4530	5	Q8I2K6	Q8i2k6 plasmodium
618	16	1.2	4638	5	Q8IK96	Q8ik96 plasmodium
619	16	1.2	4662	5	Q8ILR9	Q8ilr9 plasmodium
620	16	1.2	4894	5	Q8IBJ2	Q8ibj2 plasmodium
621	16	1.2	4981	5	O77372	O77372 plasmodium
622	16	1.2	5251	5	Q8IID4	Q8iid4 plasmodium
623	16	1.2	5439	5	Q8I586	Q8i586 plasmodium
624	16	1.2	5561	5	Q8ILC9	Q8ilc9 plasmodium
625	16	1.2	5910	5	Q8IAP1	Q8iap1 plasmodium
626	16	1.2	6077	5	Q8ICS6	Q8ics6 plasmodium
627	16	1.2	6761	5	Q8IC77	Q8ic77 plasmodium
628	16	1.2	9341	5	Q8I3N9	Q8i3n9 plasmodium
629	15	1.2	125	12	Q9PWY7	Q9pwy7 shope fibro
630	15	1.2	389	5	Q8IC70	Q8ic70 plasmodium
631	15	1.2	461	6	Q95K23	Q95k23 macaca fasc
632	15	1.2	481	5	Q8SSR3	Q8ssr3 dictyosteli
633	15	1.2	481	5	Q8I3L2	Q8i3l2 plasmodium
634	15	1.2	482	5	Q8ICP2	Q8icp2 plasmodium
635	15	1.2	515	5	O15912	O15912 dictyosteli
636	15	1.2	563	5	Q8I2B2	Q8i2b2 plasmodium
637	15	1.2	593	5	Q8I2U6	Q8i2u6 plasmodium
638	15	1.2	602	5	Q8MMM3	Q8mmm3 dictyosteli
639	15	1.2	618	5	Q8IDT5	Q8idt5 plasmodium
640	15	1.2	626	5	Q9W1U4	Q9w1u4 drosophila
641	15	1.2	631	5	Q8I1U1	Q8i1u1 plasmodium
642	15	1.2	655	5	Q8I5P8	Q8i5p8 plasmodium
643	15	1.2	670	6	Q95LT1	Q95lt1 macaca fasc
644	15	1.2	670	6	Q8WNT8	Q8wnt8 macaca fasc
645	15	1.2	673	4	Q96MA2	Q96ma2 homo sapien
646	15	1.2	673	4	Q8N4N8	Q8n4n8 homo sapien
647	15	1.2	684	10	Q940Y8	Q940y8 arabidopsis
648	15	1.2	703	10	Q9AVP4	Q9avp4 nicotiana t
649	15	1.2	706	10	Q9LW81	Q9lw81 arabidopsis
650	15	1.2	707	10	Q9FPA5	Q9fpa5 oryza sativ
651	15	1.2	725	5	Q8T124	Q8t124 dictyosteli
652	15	1.2	763	5	Q8I3P2	Q8i3p2 plasmodium
653	15	1.2	793	5	Q8SSR7	Q8ssr7 dictyosteli
654	15	1.2	800	10	Q94GW1	Q94gw1 oryza sativ
655	15	1.2	816	5	Q8T2N6	Q8t2n6 dictyosteli
656	15	1.2	817	5	Q964V6	Q964v6 dictyosteli
657	15	1.2	825	5	Q8I2K4	Q8i2k4 plasmodium
658	15	1.2	941	5	Q8IAU1	Q8iau1 plasmodium
659	15	1.2	961	5	Q8IC12	Q8ic12 plasmodium
660	15	1.2	964	5	Q8IIG7	Q8iig7 plasmodium
661	15	1.2	1025	5	Q8ICI3	Q8ici3 plasmodium
662	15	1.2	1055	5	Q8MNI5	Q8mni5 dictyosteli
663	15	1.2	1104	5	Q8I5E2	Q8i5e2 plasmodium
664	15	1.2	1139	5	Q8IJS7	Q8ijs7 plasmodium
665	15	1.2	1181	5	O97260	O97260 plasmodium
666	15	1.2	1195	5	Q8I351	Q8i351 plasmodium
667	15	1.2	1313	5	Q8IIT7	Q8iit7 plasmodium
668	15	1.2	1328	5	Q8IDY9	Q8idy9 plasmodium
669	15	1.2	1337	5	Q9Y008	Q9y008 plasmodium
670	15	1.2	1379	5	Q8I2Z2	Q8i2z2 plasmodium
671	15	1.2	1438	5	Q8IJR8	Q8ijr8 plasmodium
672	15	1.2	1461	5	Q8I5H0	Q8i5h0 plasmodium
673	15	1.2	1490	5	Q8I416	Q8i416 plasmodium

674	15	1.2	1559	5	Q8I406	Q8i406 plasmodium
675	15	1.2	1605	5	Q8IBL3	Q8ibl3 plasmodium
676	15	1.2	1655	5	Q24754	Q24754 drosophila
677	15	1.2	1731	5	Q8IBQ8	Q8ibq8 plasmodium
678	15	1.2	1813	5	Q8ID65	Q8id65 plasmodium
679	15	1.2	2041	5	Q8IBK6	Q8ibk6 plasmodium
680	15	1.2	2075	5	Q8IHR5	Q8ihr5 plasmodium
681	15	1.2	2249	5	Q8I1Y8	Q8i1y8 plasmodium
682	15	1.2	2539	5	O96157	O96157 plasmodium
683	15	1.2	2577	5	Q8I531	Q8i531 plasmodium
684	15	1.2	2706	5	O97292	O97292 plasmodium
685	15	1.2	2723	5	Q8I5X3	Q8i5x3 plasmodium
686	15	1.2	2820	5	Q8IM32	Q8im32 plasmodium
687	15	1.2	3183	5	Q8IBJ1	Q8ibj1 plasmodium
688	15	1.2	3322	5	Q8IKL0	Q8ikl0 plasmodium
689	15	1.2	3394	5	O77384	O77384 plasmodium
690	15	1.2	4044	5	Q8IDD4	Q8idd4 plasmodium
691	15	1.2	4230	5	Q8IC31	Q8ic31 plasmodium
692	15	1.2	4261	5	Q8IFP4	Q8ifp4 plasmodium
693	15	1.2	5415	5	Q8IE74	Q8ie74 plasmodium
694	15	1.2	5704	5	Q8IAT1	Q8iat1 plasmodium
695	15	1.2	6088	5	Q8IEN1	Q8ien1 plasmodium
696	14	1.1	105	5	Q8MXM8	Q8mxm8 dictyosteli
697	14	1.1	140	5	Q8IFM7	Q8ifm7 plasmodium
698	14	1.1	150	13	Q9PSH8	Q9psh8 xenopus lae
699	14	1.1	157	10	Q9AVD1	Q9avd1 nicotiana t
700	14	1.1	157	10	Q9AVD2	Q9avd2 nicotiana t
701	14	1.1	158	10	Q9FQL8	Q9fql8 cyanophora
702	14	1.1	168	11	O35068	O35068 mus musculu
703	14	1.1	205	5	Q19307	Q19307 caenorhabdi
704	14	1.1	312	10	Q8S9D7	Q8s9d7 equisetum a
705	14	1.1	347	10	Q93XF8	Q93xf8 zea mays (m
706	14	1.1	354	5	Q9BMB8	Q9bmb8 lymantria d
707	14	1.1	354	10	Q93XF5	Q93xf5 zea mays (m
708	14	1.1	375	10	O48527	O48527 arabidopsis
709	14	1.1	385	5	Q8IFP5	Q8ifp5 plasmodium
710	14	1.1	393	5	Q8T1P0	Q8t1p0 dictyosteli
711	14	1.1	407	10	Q93XF9	Q93xf9 zea mays (m
712	14	1.1	416	5	Q03752	Q03752 plasmodium
713	14	1.1	421	5	Q8I4V3	Q8i4v3 plasmodium
714	14	1.1	433	5	Q8ILH7	Q8ilh7 plasmodium
715	14	1.1	476	5	Q8IFN3	Q8ifn3 plasmodium
716	14	1.1	477	5	Q8MN77	Q8mn77 dictyosteli
717	14	1.1	495	5	Q8SUH1	Q8suh1 encephalito
718	14	1.1	521	5	Q8IDJ0	Q8idj0 plasmodium
719	14	1.1	548	5	Q8IL96	Q8il96 plasmodium
720	14	1.1	551	5	O76494	O76494 dictyosteli
721	14	1.1	555	5	Q8IJ43	Q8ij43 plasmodium
722	14	1.1	558	5	Q8I5T1	Q8i5t1 plasmodium
723	14	1.1	560	5	Q8I715	Q8i715 plasmodium
724	14	1.1	560	5	Q8I516	Q8i516 plasmodium
725	14	1.1	576	5	Q8ID42	Q8id42 plasmodium
726	14	1.1	587	5	Q8T2T9	Q8t2t9 dictyosteli
727	14	1.1	601	3	Q9C1I3	Q9c1i3 candida alb
728	14	1.1	625	5	Q8ICK3	Q8ick3 plasmodium
729	14	1.1	627	5	Q8ILX9	Q8ilx9 plasmodium
730	14	1.1	642	10	Q8S6Q6	Q8s6q6 oryza sativ
731	14	1.1	652	5	O77168	O77168 apis mellif
732	14	1.1	668	5	O60964	O60964 leishmania
733	14	1.1	677	5	Q8IC73	Q8ic73 plasmodium
734	14	1.1	678	13	Q8UUL1	Q8uul1 oryzias lat
735	14	1.1	682	5	Q8II98	Q8ii98 plasmodium
736	14	1.1	687	5	Q8IKE2	Q8ike2 plasmodium
737	14	1.1	693	5	O96680	O96680 drosophila
738	14	1.1	699	5	O15816	O15816 dictyosteli
739	14	1.1	730	5	Q8T674	Q8t674 dictyosteli
740	14	1.1	733	10	Q8L491	Q8l491 oryza sativ
741	14	1.1	767	10	O22240	O22240 arabidopsis
742	14	1.1	773	13	P79805	P79805 morone saxa
743	14	1.1	777	10	Q9M0X6	Q9m0x6 arabidopsis
744	14	1.1	794	5	Q8II27	Q8ii27 plasmodium
745	14	1.1	801	10	Q9LG54	Q9lg54 oryza sativ
746	14	1.1	810	10	Q8L4J3	Q8l4j3 oryza sativ

747	14	1.1	853	5	Q8ICM6	Q8icm6 plasmodium
748	14	1.1	867	5	Q8I3W5	Q8i3w5 plasmodium
749	14	1.1	897	10	Q9SS42	Q9ss42 arabidopsis
750	14	1.1	949	5	O77386	O77386 plasmodium
751	14	1.1	967	10	Q9FHD2	Q9fhd2 arabidopsis
752	14	1.1	974	5	Q8IED7	Q8ied7 plasmodium
753	14	1.1	982	5	Q8IDQ6	Q8idq6 plasmodium
754	14	1.1	987	10	Q9C9A8	Q9c9a8 arabidopsis
755	14	1.1	992	5	Q8ILF1	Q8ilf1 plasmodium
756	14	1.1	1011	5	Q8IAL6	Q8ial6 plasmodium
757	14	1.1	1031	5	Q8I564	Q8i564 plasmodium
758	14	1.1	1050	10	Q9C9S6	Q9c9s6 arabidopsis
759	14	1.1	1056	10	Q9CAC9	Q9cac9 arabidopsis
760	14	1.1	1078	5	Q8I2V5	Q8i2v5 plasmodium
761	14	1.1	1087	10	Q8W1Y3	Q8w1y3 arabidopsis
762	14	1.1	1093	5	Q8I5R6	Q8i5r6 plasmodium
763	14	1.1	1109	10	Q9SH47	Q9sh47 arabidopsis
764	14	1.1	1162	10	Q9LPQ7	Q9lpq7 arabidopsis
765	14	1.1	1173	5	Q8IEM5	Q8iem5 plasmodium
766	14	1.1	1232	5	Q8T821	Q8t821 dictyosteli
767	14	1.1	1241	5	Q8MNI0	Q8mni0 dictyosteli
768	14	1.1	1245	5	O96195	O96195 plasmodium
769	14	1.1	1267	5	Q8ILD1	Q8ild1 plasmodium
770	14	1.1	1268	10	Q9LUG0	Q9lug0 arabidopsis
771	14	1.1	1288	5	Q8IDE5	Q8ide5 plasmodium
772	14	1.1	1292	10	Q9LDN0	Q9ldn0 arabidopsis
773	14	1.1	1313	10	Q8L7Y8	Q8l7y8 arabidopsis
774	14	1.1	1315	5	Q9BLV2	Q9blv2 leishmania
775	14	1.1	1352	5	Q8II55	Q8ii55 plasmodium
776	14	1.1	1361	5	Q8IBV5	Q8ibv5 plasmodium
777	14	1.1	1444	5	Q8ICB9	Q8icb9 plasmodium
778	14	1.1	1449	5	Q8ILP9	Q8ilp9 plasmodium
779	14	1.1	1465	5	Q8ST07	Q8st07 dictyosteli
780	14	1.1	1467	5	Q8IAN5	Q8ian5 plasmodium
781	14	1.1	1483	5	Q8IIQ7	Q8iiq7 plasmodium
782	14	1.1	1499	13	Q9DDA6	Q9dda6 xenopus lae
783	14	1.1	1585	5	Q8ILE0	Q8ile0 plasmodium
784	14	1.1	1662	10	O23274	O23274 arabidopsis
785	14	1.1	1740	5	Q8IJK9	Q8ijk9 plasmodium
786	14	1.1	1753	5	Q8I5W3	Q8i5w3 plasmodium
787	14	1.1	1816	5	O97275	O97275 plasmodium
788	14	1.1	1838	5	Q8IDZ6	Q8idz6 plasmodium
789	14	1.1	1840	5	Q8IED3	Q8ied3 plasmodium
790	14	1.1	1891	5	O77275	O77275 drosophila
791	14	1.1	1946	5	O97291	O97291 plasmodium
792	14	1.1	1960	5	Q9U0K8	Q9u0k8 plasmodium
793	14	1.1	1967	10	Q8GVH3	Q8gvh3 oryza sativ
794	14	1.1	1991	5	Q8I515	Q8i515 plasmodium
795	14	1.1	1997	5	Q8IAP3	Q8iap3 plasmodium
796	14	1.1	2010	5	O96171	O96171 plasmodium
797	14	1.1	2031	5	Q8I1R4	Q8i1r4 plasmodium
798	14	1.1	2108	5	Q8I5C7	Q8i5c7 plasmodium
799	14	1.1	2158	10	Q9LUT5	Q9lut5 arabidopsis
800	14	1.1	2273	5	Q9U141	Q9u141 leishmania
801	14	1.1	2290	5	Q8IJI4	Q8iji4 plasmodium
802	14	1.1	2371	5	Q8IIP2	Q8iip2 plasmodium
803	14	1.1	2548	5	Q8IE95	Q8ie95 plasmodium
804	14	1.1	2558	5	Q8I5J1	Q8i5j1 plasmodium
805	14	1.1	2698	5	Q8I3F7	Q8i3f7 plasmodium
806	14	1.1	3069	5	Q8I350	Q8i350 plasmodium
807	14	1.1	3270	5	Q8IDB2	Q8idb2 plasmodium
808	14	1.1	3347	5	Q8IKV0	Q8ikv0 plasmodium
809	14	1.1	5633	5	Q8I454	Q8i454 plasmodium
810	14	1.1	5922	5	Q8IIN2	Q8iin2 plasmodium
811	14	1.1	7149	5	Q8IM09	Q8im09 plasmodium
812	13	1.0	138	5	Q8MPB1	Q8mpb1 dictyosteli
813	13	1.0	169	11	O35061	O35061 mus musculu
814	13	1.0	183	3	Q96UG3	Q96ug3 kluyveromyc
815	13	1.0	217	5	Q8IFN2	Q8ifn2 plasmodium
816	13	1.0	277	5	Q9NK49	Q9nk49 drosophila
817	13	1.0	291	5	Q8T843	Q8t843 dictyosteli
818	13	1.0	358	5	Q8MPB0	Q8mpb0 dictyosteli
819	13	1.0	367	10	Q9M2P9	Q9m2p9 arabidopsis

820	13	1.0	374	10	Q94AJ2	Q94aj2 arabidopsis
821	13	1.0	392	10	Q9ZRB9	Q9zrb9 lycopersico
822	13	1.0	407	5	Q8IE15	Q8ie15 plasmodium
823	13	1.0	411	5	Q8IDX4	Q8idx4 plasmodium
824	13	1.0	420	5	Q8SRB7	Q8srb7 encephalito
825	13	1.0	423	5	Q8IBK3	Q8ibk3 plasmodium
826	13	1.0	430	10	Q93XF6	Q93xf6 zea mays (m
827	13	1.0	439	5	Q969B0	Q969b0 giardia lam
828	13	1.0	442	5	Q8ILE4	Q8ile4 plasmodium
829	13	1.0	449	5	Q8T2E4	Q8t2e4 dictyosteli
830	13	1.0	498	5	Q9BMG7	Q9bmg7 plasmodium
831	13	1.0	507	5	Q8MNM5	Q8mnm5 dictyosteli
832	13	1.0	511	5	Q8MXL4	Q8mxl4 dictyosteli
833	13	1.0	517	5	Q8IJP2	Q8ijp2 plasmodium
834	13	1.0	575	5	Q8IM39	Q8im39 plasmodium
835	13	1.0	592	5	Q8IDT9	Q8idt9 plasmodium
836	13	1.0	597	5	Q8I476	Q8i476 plasmodium
837	13	1.0	601	5	Q8IJJ8	Q8ijj8 plasmodium
838	13	1.0	604	5	Q8IS16	Q8is16 dictyosteli
839	13	1.0	608	5	O77331	O77331 plasmodium
840	13	1.0	647	5	Q9NH62	Q9nh62 schistosoma
841	13	1.0	649	5	Q8T1Q9	Q8t1q9 dictyosteli
842	13	1.0	674	5	Q8IB50	Q8ib50 plasmodium
843	13	1.0	679	5	Q9W3E3	Q9w3e3 drosophila
844	13	1.0	684	5	Q8I2M7	Q8i2m7 plasmodium
845	13	1.0	707	5	O00910	O00910 dictyosteli
846	13	1.0	717	5	Q9VW27	Q9vw27 drosophila
847	13	1.0	734	5	Q8I1Z1	Q8i1z1 plasmodium
848	13	1.0	742	5	Q8I5E7	Q8i5e7 plasmodium
849	13	1.0	743	5	Q8ILV4	Q8ilv4 plasmodium
850	13	1.0	771	5	Q9XZJ3	Q9xzj3 dictyosteli
851	13	1.0	786	5	Q8T1B7	Q8t1b7 dictyosteli
852	13	1.0	794	10	Q940B8	Q940b8 arabidopsis
853	13	1.0	796	5	Q8IEN5	Q8ien5 plasmodium
854	13	1.0	799	10	Q9LUS1	Q9lus1 arabidopsis
855	13	1.0	818	5	Q8SS10	Q8ss10 encephalito
856	13	1.0	838	5	Q8ILS1	Q8ils1 plasmodium
857	13	1.0	838	5	Q8IJB9	Q8ijb9 plasmodium
858	13	1.0	842	5	Q8I5J8	Q8i5j8 plasmodium
859	13	1.0	857	5	Q8IBF8	Q8ibf8 plasmodium
860	13	1.0	863	5	Q8ICT6	Q8ict6 plasmodium
861	13	1.0	865	6	Q95LL1	Q95ll1 macaca fasc
862	13	1.0	899	5	Q8MMR2	Q8mmr2 dictyosteli
863	13	1.0	904	3	Q04195	Q04195 saccharomyc
864	13	1.0	905	5	Q8ID72	Q8id72 plasmodium
865	13	1.0	936	5	Q8I2A7	Q8i2a7 plasmodium
866	13	1.0	937	5	Q8MP59	Q8mp59 dictyosteli
867	13	1.0	938	5	Q8IHY0	Q8ihy0 plasmodium
868	13	1.0	954	5	Q9XYL1	Q9xyl1 dictyosteli
869	13	1.0	960	5	Q8IEC6	Q8iec6 plasmodium
870	13	1.0	975	5	Q8IAY8	Q8iay8 plasmodium
871	13	1.0	1011	5	Q8IEP4	Q8iep4 plasmodium
872	13	1.0	1016	5	Q8I583	Q8i583 plasmodium
873	13	1.0	1029	5	Q8IKV7	Q8ikv7 plasmodium
874	13	1.0	1032	10	Q9FIJ9	Q9fij9 arabidopsis
875	13	1.0	1033	5	Q8IBB8	Q8ibb8 plasmodium
876	13	1.0	1035	10	Q8GS71	Q8gs71 arabidopsis
877	13	1.0	1044	5	Q9GT38	Q9gt38 plasmodium
878	13	1.0	1044	5	Q8I570	Q8i570 plasmodium
879	13	1.0	1048	5	Q9VKW4	Q9vkw4 drosophila
880	13	1.0	1048	5	Q9NGN7	Q9ngn7 drosophila
881	13	1.0	1051	10	Q94LW7	Q94lw7 arabidopsis
882	13	1.0	1063	5	Q8ICV5	Q8icv5 plasmodium
883	13	1.0	1065	5	Q24029	Q24029 drosophila
884	13	1.0	1065	5	Q8IP21	Q8ip21 drosophila
885	13	1.0	1072	5	Q9VJK6	Q9vjk6 drosophila
886	13	1.0	1072	5	Q24031	Q24031 drosophila
887	13	1.0	1072	5	Q24027	Q24027 drosophila
888	13	1.0	1073	5	Q8IJW5	Q8ijw5 plasmodium
889	13	1.0	1074	5	Q24028	Q24028 drosophila
890	13	1.0	1074	5	Q8IP20	Q8ip20 drosophila
891	13	1.0	1075	10	Q9SNE3	Q9sne3 arabidopsis
892	13	1.0	1076	5	Q8IL74	Q8il74 plasmodium
893	13	1.0	1081	5	Q24030	Q24030 drosophila
894	13	1.0	1081	5	Q8IP19	Q8ip19 drosophila
895	13	1.0	1083	5	O45935	O45935 caenorhabdi
896	13	1.0	1111	5	Q8IJA2	Q8ija2 plasmodium
897	13	1.0	1139	5	O97237	O97237 plasmodium
898	13	1.0	1148	5	Q8MYD8	Q8myd8 dictyosteli
899	13	1.0	1158	5	Q8IU40	Q8iu40 dictyosteli
900	13	1.0	1163	5	Q8MP92	Q8mp92 dictyosteli
901	13	1.0	1172	5	Q8I5E4	Q8i5e4 plasmodium
902	13	1.0	1191	10	Q8GZU1	Q8gzu1 lycopersico
903	13	1.0	1212	5	Q9XZ29	Q9xz29 drosophila
904	13	1.0	1215	5	Q8I1Q6	Q8i1q6 plasmodium
905	13	1.0	1225	5	Q8ILF3	Q8ilf3 plasmodium
906	13	1.0	1235	5	Q8IB22	Q8ib22 plasmodium
907	13	1.0	1280	5	Q8ICQ7	Q8icq7 plasmodium
908	13	1.0	1281	5	O15802	O15802 plasmodium
909	13	1.0	1281	5	Q8IBZ5	Q8ibz5 plasmodium
910	13	1.0	1299	5	Q8IJQ0	Q8ijq0 plasmodium
911	13	1.0	1316	5	Q8IAS8	Q8ias8 plasmodium
912	13	1.0	1318	5	Q8IDI3	Q8idi3 plasmodium
913	13	1.0	1335	5	O00886	O00886 dictyosteli
914	13	1.0	1335	10	Q9FME7	Q9fme7 arabidopsis
915	13	1.0	1341	5	Q8IM28	Q8im28 plasmodium
916	13	1.0	1342	5	O97364	O97364 plasmodium
917	13	1.0	1384	5	Q8I3B3	Q8i3b3 plasmodium
918	13	1.0	1401	5	Q8I5D4	Q8i5d4 plasmodium
919	13	1.0	1404	5	Q8IKT7	Q8ikt7 plasmodium
920	13	1.0	1411	5	Q8I277	Q8i277 plasmodium
921	13	1.0	1495	5	Q8MML8	Q8mml8 dictyosteli
922	13	1.0	1542	5	Q8I2T2	Q8i2t2 plasmodium
923	13	1.0	1550	5	Q8IDI7	Q8idi7 plasmodium
924	13	1.0	1554	5	Q8IJY4	Q8ijy4 plasmodium
925	13	1.0	1575	5	Q8SSN3	Q8ssn3 dictyosteli
926	13	1.0	1594	5	Q8I2N7	Q8i2n7 plasmodium
927	13	1.0	1701	5	Q8II50	Q8ii50 plasmodium
928	13	1.0	1706	5	Q8I2W9	Q8i2w9 plasmodium
929	13	1.0	1737	5	Q8IEH8	Q8ieh8 plasmodium
930	13	1.0	1754	5	Q8T1C5	Q8t1c5 dictyosteli
931	13	1.0	1790	5	Q9BK49	Q9bk49 drosophila
932	13	1.0	1816	5	Q8IJL6	Q8ijl6 plasmodium
933	13	1.0	1830	5	Q8IIR3	Q8iir3 plasmodium
934	13	1.0	1893	5	Q9W4J1	Q9w4j1 drosophila
935	13	1.0	1903	5	Q9U0N7	Q9u0n7 plasmodium
936	13	1.0	1920	5	O46205	O46205 drosophila
937	13	1.0	2084	5	Q8T129	Q8t129 dictyosteli
938	13	1.0	2121	5	Q9GYC1	Q9gyc1 leishmania
939	13	1.0	2269	5	O77360	O77360 plasmodium
940	13	1.0	2310	5	Q8IBZ1	Q8ibz1 plasmodium
941	13	1.0	2558	5	Q8IKF6	Q8ikf6 plasmodium
942	13	1.0	2573	5	O96185	O96185 plasmodium
943	13	1.0	2653	5	Q8IHX2	Q8ihx2 plasmodium
944	13	1.0	2657	5	Q8ILG1	Q8ilg1 plasmodium
945	13	1.0	2691	5	Q8ILS2	Q8ils2 plasmodium
946	13	1.0	2705	5	Q8I5L0	Q8i5l0 plasmodium
947	13	1.0	3001	5	Q8IL08	Q8il08 plasmodium
948	13	1.0	3063	5	Q8I284	Q8i284 plasmodium
949	13	1.0	3377	5	Q8IEH5	Q8ieh5 plasmodium
950	13	1.0	3429	5	Q8IBP1	Q8ibp1 plasmodium
951	13	1.0	3535	5	Q8IC29	Q8ic29 plasmodium
952	13	1.0	3574	5	Q8IDA8	Q8ida8 plasmodium
953	13	1.0	3610	5	Q968T7	Q968t7 plasmodium
954	13	1.0	3628	5	Q968Y0	Q968y0 plasmodium
955	13	1.0	3704	5	Q8IKY8	Q8iky8 plasmodium
956	13	1.0	5767	5	Q8I525	Q8i525 plasmodium
957	13	1.0	5890	5	Q8IK84	Q8ik84 plasmodium
958	13	1.0	5987	5	Q8IDN3	Q8idn3 plasmodium
959	13	1.0	6118	5	Q8I396	Q8i396 plasmodium
960	13	1.0	6473	5	Q8IKH9	Q8ikh9 plasmodium
961	12	0.9	74	5	Q9GVB2	Q9gvb2 plasmodium
962	12	0.9	106	10	Q8LT11	Q8lt11 lycopersico
963	12	0.9	108	5	Q8T211	Q8t211 dictyosteli
964	12	0.9	125	5	Q8T2H7	Q8t2h7 dictyosteli
965	12	0.9	155	11	Q99PT6	Q99pt6 mus musculu

```
 966     12    0.9    196  11  Q8C2X4          Q8c2x4 mus musculu
 967     12    0.9    239  5   Q25772          Q25772 plasmodium
 968     12    0.9    246  5   Q9U4U4          Q9u4u4 cryptospori
 969     12    0.9    263  5   Q9W2T4          Q9w2t4 drosophila
 970     12    0.9    278  15  Q90UP8          Q90up8 equine infe
 971     12    0.9    301  5   Q8IE30          Q8ie30 plasmodium
 972     12    0.9    306  5   Q9VM35          Q9vm35 drosophila
 973     12    0.9    319  5   Q94499          Q94499 dictyosteli
 974     12    0.9    325  5   Q94492          Q94492 dictyosteli
 975     12    0.9    327  5   Q8MXP9          Q8mxp9 dictyosteli
 976     12    0.9    334  5   Q8T2M7          Q8t2m7 dictyosteli
 977     12    0.9    334  5   Q8I2B5          Q8i2b5 plasmodium
 978     12    0.9    373  10  Q8L841          Q8l841 arabidopsis
 979     12    0.9    387  3   Q8X0Z8          Q8x0z8 candida alb
 980     12    0.9    388  5   Q8IIM8          Q8iim8 plasmodium
 981     12    0.9    389  13  Q9PTK9          Q9ptk9 brachydanio
 982     12    0.9    399  8   Q9B6E2          Q9b6e2 yarrowia li
 983     12    0.9    403  5   Q8IBV2          Q8ibv2 plasmodium
 984     12    0.9    404  8   Q95946          Q95946 saccharomyc
 985     12    0.9    405  10  Q93XF2          Q93xf2 zea mays (m
 986     12    0.9    420  5   O76853          O76853 dictyosteli
 987     12    0.9    427  5   Q9VIJ6          Q9vij6 drosophila
 988     12    0.9    431  3   Q12124          Q12124 saccharomyc
 989     12    0.9    451  5   Q95ZC1          Q95zc1 leishmania
 990     12    0.9    468  4   Q8N5N1          Q8n5n1 homo sapien
 991     12    0.9    522  5   Q8MN27          Q8mn27 dictyosteli
 992     12    0.9    551  5   Q8ILY2          Q8ily2 plasmodium
 993     12    0.9    563  5   Q8T661          Q8t661 dictyosteli
 994     12    0.9    578  5   Q8I387          Q8i387 plasmodium
 995     12    0.9    579  5   O76734          O76734 dictyosteli
 996     12    0.9    588  5   Q8MMM8          Q8mmm8 dictyosteli
 997     12    0.9    603  10  Q9C6K3          Q9c6k3 arabidopsis
 998     12    0.9    611  5   Q8I449          Q8i449 plasmodium
 999     12    0.9    616  5   Q8I5Q2          Q8i5q2 plasmodium
1000     12    0.9    640  5   Q8I6Z8          Q8i6z8 plasmodium
```

ALIGNMENTS

```
RESULT 1
Q8I4Y0
ID   Q8I4Y0      PRELIMINARY;      PRT;  1351 AA.
AC   Q8I4Y0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Kinesin-like protein, putative.
GN   PFL2165W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014851; AAN36517.1; -.
SQ   SEQUENCE   1351 AA;  156007 MW;  8B69594BFD73C788 CRC64;

  Query Match             100.0%;  Score 1288;  DB 5;  Length 1351;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 123

Qy         61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 183

Qy        121 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 243

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 303

Qy        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 363

Qy        301 GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGS 423

Qy        361 EMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 EMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKY 483

Qy        421 SSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNDNDNDNNNDNNNNNNNDSS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 SSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNDNDNDNNNDNNNNNNNDSS 543

Qy        481 SMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 SMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 603

Qy        541 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 663

Qy        601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 723

Qy        661 NGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 NGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNK 783

Qy        721 NNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 NNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNS 843

Qy        781 HNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIH 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 HNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIH 903

Qy        841 GCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 GCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNE 963

Qy        901 MKNNEMKDNEMKDNHIKSNNNNSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKN 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 MKNNEMKDNEMKDNHIKSNNNNSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKN 1023

Qy        961 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLD 1083

Qy       1021 DQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIK 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 DQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIK 1143
```

```
Qy       1081 NFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYIN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 NFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYIN 1203

Qy       1141 EDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 EDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN 1263

Qy       1201 NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSN 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1264 NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSN 1323

Qy       1261 NGDIILLNKKLVQDNIKNSMDHNNIHKK 1288
              ||||||||||||||||||||||||||||
Db       1324 NGDIILLNKKLVQDNIKNSMDHNNIHKK 1351


RESULT 2
Q8T224
ID   Q8T224      PRELIMINARY;      PRT;  1723 AA.
AC   Q8T224;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 188.9 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116030; AAL92981.1; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00036; efhand; 1.
DR   Pfam; PF00046; homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1723 AA;  188915 MW;  F68A3B702B3FEB95 CRC64;

  Query Match             1.5%;  Score 19;  DB 5;  Length 1723;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-09;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        739 NITNNNNNNNNNNNNNNNN 757
              |||||||||||||||||||
Db        544 NITNNNNNNNNNNNNNNNN 562


RESULT 3
Q8MN71
ID   Q8MN71      PRELIMINARY;      PRT;   348 AA.
AC   Q8MN71;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117079; AAM43722.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   348 AA;  39051 MW;  9A3DDDC4941FF548 CRC64;

  Query Match             1.4%;  Score 18;  DB 5;  Length 348;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-08;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNID 759
              ||||||||||||||||||
Db        164 NNNNNNNNNNNNNNNNID 181


RESULT 4
Q8T832
ID   Q8T832      PRELIMINARY;      PRT;   439 AA.
AC   Q8T832;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 51.5 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115581; AAL92229.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   439 AA;  51503 MW;  D8EC69EA046B0C2E CRC64;

  Query Match             1.4%;  Score 18;  DB 5;  Length 439;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-08;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        740 ITNNNNNNNNNNNNNNNN 757
              ||||||||||||||||||
Db        358 ITNNNNNNNNNNNNNNNN 375


RESULT 5
Q8MN10
ID   Q8MN10      PRELIMINARY;      PRT;   699 AA.
AC   Q8MN10;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117177; AAM43662.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   699 AA;  79820 MW;  5A1D3995C439AE4D CRC64;

  Query Match             1.4%;  Score 18;  DB 5;  Length 699;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-08;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        742 NNNNNNNNNNNNNNNNID 759
              ||||||||||||||||||
Db        111 NNNNNNNNNNNNNNNNID 128


RESULT 6
Q8T686
ID   Q8T686      PRELIMINARY;      PRT;   815 AA.
AC   Q8T686;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   ABC transporter AbcG7.
GN   ABCG7.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ax4;
RA   Anjard.C., Loomis W.F.;
RT   "Evolution of the ABC transporters of Dictyostelium.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF482386; AAL91492.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding.
SQ   SEQUENCE   815 AA;  92419 MW;  5665336EFED31C95 CRC64;

  Query Match             1.4%;  Score 18;  DB 5;  Length 815;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-08;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNID 759
              ||||||||||||||||||
Db        102 NNNNNNNNNNNNNNNNID 119


RESULT 7
Q8T1J6
ID   Q8T1J6      PRELIMINARY;      PRT;  1518 AA.
AC   Q8T1J6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 174.9 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116425; AAL99110.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   Hypothetical protein; ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE   1518 AA;  174934 MW;  8FE54FB086031500 CRC64;

  Query Match             1.4%;  Score 18;  DB 5;  Length 1518;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-08;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        740 ITNNNNNNNNNNNNNNNN 757
              ||||||||||||||||||
Db       1060 ITNNNNNNNNNNNNNNNN 1077


RESULT 8
Q8SSS8
ID   Q8SSS8      PRELIMINARY;      PRT;  1887 AA.
AC   Q8SSS8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Y54G2A.A protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AC116029; AAL96739.1; -.
DR   InterPro; IPR000569; HECT_domain.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR000408; Reg_chr_condens.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00632; HECT; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00415; RCC1; 3.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00119; HECTc; 1.
DR   PROSITE; PS50237; HECT; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS00626; RCC1_2; 2.
DR   PROSITE; PS50012; RCC1_3; 6.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   1887 AA;  213010 MW;  6E0D63ED31793F65 CRC64;

  Query Match             1.4%;  Score 18;  DB 5;  Length 1887;
  Best Local Similarity   100.0%;  Pred. No. 6.2e-08;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNID 759
              ||||||||||||||||||
Db       1052 NNNNNNNNNNNNNNNNID 1069


RESULT 9
Q95ZG5
ID   Q95ZG5      PRELIMINARY;      PRT;  2994 AA.
AC   Q95ZG5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative RNaseIII.
GN   DRNA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Martens H., Novotny J., Oberstrass J., Steck T.L., Postlethwait P.,
RA   Nellen W.;
```

```
RT   "RNAi in Dictyostelium: developmental regulation and the role of RdRPs
RT   and dsRNase.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ314911; CAC41976.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000999; RNase_3.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF04243; DUF425; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00636; Ribonuclease_3; 2.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00535; RIBOc; 2.
DR   PROSITE; PS00517; RNASE_3_1; 2.
DR   PROSITE; PS50142; RNASE_3_2; 2.
KW   ATP-binding; Helicase; Hydrolase.
SQ   SEQUENCE   2994 AA;  348080 MW;  E88B2E33AFBACCD4 CRC64;

  Query Match             1.4%;  Score 18;  DB 5;  Length 2994;
  Best Local Similarity   100.0%;  Pred. No. 9e-08;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        740 ITNNNNNNNNNNNNNNNN 757
              ||||||||||||||||||
Db       1386 ITNNNNNNNNNNNNNNNN 1403


RESULT 10
Q9NA13
ID   Q9NA13      PRELIMINARY;      PRT;  3177 AA.
AC   Q9NA13;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Inositol 1,4,5-trisphosphate receptor-like protein.
GN   IPLA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ax2;
RX   MEDLINE=20428450; PubMed=10970875;
RA   Traynor D., Milne J.L.S., Insall R.H., Kay R.R.;
RT   "Ca2 signalling is not required for chemotaxis in Dictyostelium.";
RL   EMBO J. 19:4846-4854(2000).
DR   EMBL; AJ277590; CAB97127.1; -.
DR   InterPro; IPR000699; Ca-rel_channel.
DR   InterPro; IPR001682; Ca/Na_pore.
DR   InterPro; IPR000493; InsP3_receptor.
DR   Pfam; PF01365; RYDR_ITPR; 2.
DR   PRINTS; PR00779; INSP3RECEPTR.
KW   Receptor.
SQ   SEQUENCE   3177 AA;  360526 MW;  381ECEA20A7C84A2 CRC64;

  Query Match             1.4%;  Score 18;  DB 5;  Length 3177;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-08;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNID 759
              ||||||||||||||||||
Db        593 NNNNNNNNNNNNNNNNID 610

RESULT 11
Q9U573
ID   Q9U573      PRELIMINARY;      PRT;  3619 AA.
AC   Q9U573;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Beige protein homolog.
GN   LVSA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DH1;
RX   MEDLINE=20056115; PubMed=10588668;
RA   Kwak E., Gerald N., Larochelle D.A., Vithalani K.K., Niswonger M.L.,
RA   Maready M., De Lozanne A.;
RT   "LvsA, a protein related to the mouse beige protein, is required for
RT   cytokinesis in Dictyostelium.";
RL   Mol. Biol. Cell 10:4429-4439(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DH1;
RA   De Lozanne A.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF088979; AAD52096.2; -.
DR   InterPro; IPR000409; Beige_BEACH.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF02138; Beach; 1.
DR   Pfam; PF00400; WD40; 5.
DR   ProDom; PD007848; Beige_BEACH; 1.
DR   ProDom; PD000018; WD40; 1.
DR   SMART; SM00320; WD40; 4.
DR   PROSITE; PS50197; BEACH; 1.
DR   PROSITE; PS00678; WD_REPEATS_1; 1.
DR   PROSITE; PS50082; WD_REPEATS_2; 1.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Repeat; WD repeat.
SQ   SEQUENCE   3619 AA;  408069 MW;  24B74A953B2885F9 CRC64;

  Query Match             1.4%;  Score 18;  DB 5;  Length 3619;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-07;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        740 ITNNNNNNNNNNNNNNNN 757
              ||||||||||||||||||
Db       2600 ITNNNNNNNNNNNNNNNN 2617


RESULT 12
Q8MNQ5
ID   Q8MNQ5      PRELIMINARY;      PRT;   104 AA.
AC   Q8MNQ5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116961; AAM33740.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   104 AA;  12132 MW;  7694A678A730A5F4 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 104;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db         75 NNNNNNNNNNNNNNNNI 91
```

```
RESULT 13
Q8MP24
ID   Q8MP24      PRELIMINARY;      PRT;   172 AA.
AC   Q8MP24;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC123513; AAM44372.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   172 AA;  20800 MW;  CE6A456B2B3CFDCF CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 172;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db        111 TNNNNNNNNNNNNNNNN 127

RESULT 14
Q8ST26
ID   Q8ST26      PRELIMINARY;      PRT;   204 AA.
AC   Q8ST26;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   HSPC305 10/100.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115575; AAL92188.1; -.
SQ   SEQUENCE   204 AA;  23921 MW;  EFCD3733C7BA9C81 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 204;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        156 NNNNNNNNNNNNNNNNI 172

RESULT 15
Q8T1B5
ID   Q8T1B5      PRELIMINARY;      PRT;   279 AA.
AC   Q8T1B5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 31.4 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116960; AAM08485.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   279 AA;  31413 MW;  182A4849594CBD04 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db        100 TNNNNNNNNNNNNNNNN 116

RESULT 16
Q8ST23
ID   Q8ST23      PRELIMINARY;      PRT;   284 AA.
AC   Q8ST23;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   HDCKB03P 10/100.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115681; AAL92673.1; -.
SQ   SEQUENCE   284 AA;  32029 MW;  E522D2299F7D526A CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        197 NNNNNNNNNNNNNNNNI 213

RESULT 17
Q9XY02
ID   Q9XY02      PRELIMINARY;      PRT;   288 AA.
AC   Q9XY02;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PLOX5-Dj (Fragment).
GN   PLOX5-DJ.
OS   Dugesia japonica (Planarian).
OC   Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Seriata; Tricladida;
OC   Paludicola; Dugesiidae; Dugesia.
OX   NCBI_TaxID=6161;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GI;
RX   MEDLINE=99287788; PubMed=10357903;
RA   Orii H., Kato K., Umesono Y., Sakurai T., Agata K., Watanabe K.;
```

```
RT   "The planarian HOM/HOX homeobox genes (Plox) expressed along the
RT   antero-posterior axis.";
RL   Dev. Biol. 210:456-468(1999).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; AB024409; BAA77405.1; -.
DR   HSSP; P02833; 9ANT.
DR   InterPro; IPR001827; Antennapedia.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00025; ANTENNAPEDIA.
DR   PRINTS; PR00024; HOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Homeobox; Nuclear protein.
FT   NON_TER      1      1
SQ   SEQUENCE   288 AA;  33191 MW;  693C6496569B290A CRC64;

  Query Match              1.3%;  Score 17;  DB 5;  Length 288;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db        146 TNNNNNNNNNNNNNNNN 162


RESULT 18
Q8MX75
ID   Q8MX75      PRELIMINARY;      PRT;   293 AA.
AC   Q8MX75;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CP47 (Fragment).
GN   CP47.
OS   Cryptosporidium parvum.
OC   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC   Cryptosporidiidae; Cryptosporidium.
OX   NCBI_TaxID=5807;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jenkins M.C., Trout J.M., Fayer R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF384127; AAM46174.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   293 AA;  32688 MW;  7C81729D50B46E0D CRC64;

  Query Match              1.3%;  Score 17;  DB 5;  Length 293;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db         97 TNNNNNNNNNNNNNNNN 113


RESULT 19
Q8MNN2
ID   Q8MNN2      PRELIMINARY;      PRT;   300 AA.
AC   Q8MNN2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116965; AAM33218.1; -.
DR   InterPro; IPR001159; DS_RBD.
DR   Pfam; PF00035; dsrm; 1.
DR   SMART; SM00358; DSRM; 1.
DR   PROSITE; PS50137; DS_RBD; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   300 AA;  35188 MW;  C4EC727AF8C60E2D CRC64;

  Query Match              1.3%;  Score 17;  DB 5;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        185 NNNNNNNNNNNNNNNNI 201


RESULT 20
Q8ICP1
ID   Q8ICP1      PRELIMINARY;      PRT;   371 AA.
AC   Q8ICP1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   MAL6P1.298.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
RA   Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA   Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL844505; CAD50400.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   371 AA;  43460 MW;  8A81197CA8F81BC2 CRC64;

  Query Match              1.3%;  Score 17;  DB 5;  Length 371;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        227 NNNNNNNNNNNNNNNNI 243


RESULT 21
Q8IHR1
ID   Q8IHR1      PRELIMINARY;      PRT;   384 AA.
AC   Q8IHR1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PF11_0468.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
```

```
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014843; AAN36048.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   384 AA;  46253 MW;  F20DB4F1EF5089CB CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        211 TNNNNNNNNNNNNNNNN 227


RESULT 22
Q25768
ID   Q25768      PRELIMINARY;      PRT;   391 AA.
AC   Q25768;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Asparagine-rich antigen (Clone 14C1) (Fragment).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schreiber M.G., Deutsche U., Storck T., Mueller-Hill B.;
RL   Submitted (DEC-1989) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X17483; CAA35519.1; -.
FT   NON_TER       1      1
FT   NON_TER     391    391
SQ   SEQUENCE   391 AA;  45309 MW;  336596FF24F071A8 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        162 TNNNNNNNNNNNNNNNN 178


RESULT 23
Q25774
ID   Q25774      PRELIMINARY;      PRT;   419 AA.
AC   Q25774;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Asparagine-rich antigen (Clone 52C11) (Fragment).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schreiber M.G., Deutsche U., Storck T., Mueller-Hill B.;
RL   Submitted (DEC-1989) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X17489; CAA35525.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   419 AA;  47365 MW;  08DFBD456298DE31 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db         25 TNNNNNNNNNNNNNNNN 41


RESULT 24
Q8MNF6
ID   Q8MNF6      PRELIMINARY;      PRT;   426 AA.
AC   Q8MNF6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117072; AAM33165.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   426 AA;  48800 MW;  10D360C7BA3AC69A CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 426;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        401 NNNNNNNNNNNNNNNNI 417


RESULT 25
Q8T834
ID   Q8T834      PRELIMINARY;      PRT;   434 AA.
AC   Q8T834;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 49.6 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115581; AAL92226.1; -.
DR   InterPro; IPR001594; Znf_DHHC.
DR   Pfam; PF01529; zf-DHHC; 1.
DR   ProDom; PD003041; Znf_DHHC; 1.
DR   PROSITE; PS50216; ZF_DHHC; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   434 AA;  49641 MW;  256228913AA55E9A CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 434;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        288 NNNNNNNNNNNNNNNNI 304


RESULT 26
```

```
Q8I4W5
ID   Q8I4W5      PRELIMINARY;      PRT;   440 AA.
AC   Q8I4W5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFL2240W.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014851; AAN36532.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   440 AA;  51713 MW;  E2FBF153E1039566 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 440;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db        253 TNNNNNNNNNNNNNNNN 269

RESULT 27
Q8MXW8
ID   Q8MXW8      PRELIMINARY;      PRT;   455 AA.
AC   Q8MXW8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Winged helix/forkhead transcription factor (Fragment).
GN   DJFOXD.
OS   Dugesia japonica (Planarian).
OC   Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Seriata; Tricladida;
OC   Paludicola; Dugesiidae; Dugesia.
OX   NCBI_TaxID=6161;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GI;
RA   Koinuma S., Umesono Y., Watanabe K., Agata K.;
RT   "The expression of planarian brain factor homologues, DjFoxG and
RT   DjFoxD.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB091062; BAC10918.1; -.
DR   InterPro; IPR001766; TF_Fork_head.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM00339; FH; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   455 AA;  53245 MW;  1F355B9F5BEF2140 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db        357 TNNNNNNNNNNNNNNNN 373

RESULT 28
Q8IIJ6
ID   Q8IIJ6      PRELIMINARY;      PRT;   465 AA.
AC   Q8IIJ6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Ubiquitin C-terminal hydrolase, family 1, putative.
GN   PF11_0177.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=3D7;
RX   MEDLINE=22255705; PubMed=12368864;
RA   Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA   Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA   Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA   Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,
RA   Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA   Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA   McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA   Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA   Fraser C.M., Barrell B.;
RT   "Genome sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Nature 419:498-511(2002).
DR   EMBL; AE014838; AAN35761.1; -.
KW   Hydrolase.
SQ   SEQUENCE   465 AA;  54460 MW;  138A55EB774F99DD CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 465;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        277 NNNNNNNNNNNNNNNNI 293

RESULT 29
Q8SSX7
ID   Q8SSX7      PRELIMINARY;      PRT;   496 AA.
AC   Q8SSX7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein E01A2.2. 9/101.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115685; AAL92706.1; -.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   PROSITE; PS50102; RRM; 1.
SQ   SEQUENCE   496 AA;  56227 MW;  614DC0EEDE769126 CRC64;
```

```
  Query Match             1.3%;  Score 17;  DB 5;  Length 496;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db         94 NNNNNNNNNNNNNNNNI 110

RESULT 30
Q8MP37
ID   Q8MP37      PRELIMINARY;      PRT;   498 AA.
AC   Q8MP37;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117078; AAM45290.1; -.
DR   InterPro; IPR003034; SAP.
KW   Hypothetical protein.
SQ   SEQUENCE   498 AA;  57878 MW;  ADAA58F246919E9A CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 498;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        464 NNNNNNNNNNNNNNNNI 480

RESULT 31
Q8SSZ5
ID   Q8SSZ5      PRELIMINARY;      PRT;   511 AA.
AC   Q8SSZ5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein. 6/101.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116330; AAL96721.1; -.
DR   InterPro; IPR003734; DUF155.
DR   Pfam; PF02582; DUF155; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   511 AA;  58641 MW;  D1F00BB910D2B0D5 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 511;
  Best Local Similarity   100.0%;  Pred. No. 2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        178 NNNNNNNNNNNNNNNNI 194

RESULT 32
Q8T1K7
ID   Q8T1K7      PRELIMINARY;      PRT;   537 AA.
AC   Q8T1K7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 63.3 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116425; AAL99099.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   537 AA;  63290 MW;  A935AD400FE8137D CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 537;
  Best Local Similarity   100.0%;  Pred. No. 2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        244 NNNNNNNNNNNNNNNNI 260

RESULT 33
O61085
ID   O61085      PRELIMINARY;      PRT;   560 AA.
AC   O61085;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Coronin binding protein.
GN   DB10.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   de Hostos E.L., Benhayon D., Berger M., Gu W.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF051898; AAC05577.1; -.
SQ   SEQUENCE   560 AA;  61281 MW;  F417511014157229 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 560;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        454 TNNNNNNNNNNNNNNNN 470

RESULT 34
Q8MNG3
ID   Q8MNG3      PRELIMINARY;      PRT;   587 AA.
AC   Q8MNG3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
```

```
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117072; AAM33146.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   587 AA;  66683 MW;  4B960B35F9379E06 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db         59 NNNNNNNNNNNNNNNNI 75


RESULT 35
Q8MYG9
ID   Q8MYG9      PRELIMINARY;      PRT;   597 AA.
AC   Q8MYG9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Uracil-DNA glycosylase.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117073; AAM43634.1; -.
DR   InterPro; IPR002043; UDNA_glycsylse.
DR   InterPro; IPR005122; UDNA_glycsylseSF.
DR   InterPro; IPR003249; U_glycsylse_notp.
DR   Pfam; PF03167; UDG; 1.
DR   ProDom; PD001589; U_glycsylse_notp; 1.
DR   TIGRFAMs; TIGR00628; ung; 1.
SQ   SEQUENCE   597 AA;  68757 MW;  4FA01507B04D04CF CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 597;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        475 NNNNNNNNNNNNNNNNI 491


RESULT 36
Q94474
ID   Q94474      PRELIMINARY;      PRT;   608 AA.
AC   Q94474;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   ORFveg132 (Fragment).
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Loomis W.F.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U66369; AAB06761.1; -.
FT   NON_TER       1      1
FT   NON_TER     608    608
SQ   SEQUENCE   608 AA;  68205 MW;  22CF2B5AE402E2BC CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 608;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        345 NNNNNNNNNNNNNNNNI 361


RESULT 37
Q8T1X4
ID   Q8T1X4      PRELIMINARY;      PRT;   614 AA.
AC   Q8T1X4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 71.7 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116032; AAL93044.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   614 AA;  71715 MW;  2034B6D48222F622 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 614;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        106 NNNNNNNNNNNNNNNNI 122


RESULT 38
Q8MNJ9
ID   Q8MNJ9      PRELIMINARY;      PRT;   616 AA.
AC   Q8MNJ9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116984; AAM33199.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   616 AA;  71182 MW;  9940E42E65B85F0D CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 616;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-07;
```

```
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        270 NNNNNNNNNNNNNNNNI 286


RESULT 39
Q8SSY0
ID   Q8SSY0      PRELIMINARY;      PRT;   628 AA.
AC   Q8SSY0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   FMTA-like protein. 10/100.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115581; AAL92233.1; -.
DR   InterPro; IPR001466; Beta_lactamase.
DR   Pfam; PF00144; beta-lactamase; 1.
SQ   SEQUENCE   628 AA;  72065 MW;  335F256D12173915 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 628;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db        243 TNNNNNNNNNNNNNNNN 259


RESULT 40
Q8MN19
ID   Q8MN19      PRELIMINARY;      PRT;   634 AA.
AC   Q8MN19;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
DR   EMBL; AC117082; AAM34384.1; -.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000306; Znf_FYVE.
DR   Pfam; PF01363; FYVE; 1.
DR   Pfam; PF00169; PH; 1.
DR   SMART; SM00064; FYVE; 1.
DR   SMART; SM00233; PH; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50178; ZF_FYVE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   634 AA;  71127 MW;  0C35565610AC89C7 CRC64;

  Query Match             1.3%;  Score 17;  DB 5;  Length 634;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        132 NNNNNNNNNNNNNNNNI 148

Search completed: October  2, 2003, 16:35:27
Job time : 170 secs
```

```
OM protein - protein search, using sw model

Run on:         October  2, 2003, 16:35:32 ; Search time 85 Seconds
                                             (without alignments)
                                             2397.392 Million cell updates/sec

Title:          US-10-006-780-2
Perfect score:  1288
Sequence:       1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       587654 seqs, 158212981 residues

Word size :       0

Total number of hits satisfying chosen parameters:      587654

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :       Published_Applications_AA:*
                 1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                 2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                 3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                 4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                 5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                 6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                 7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                 8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                 9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                 10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                 11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                 12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                 13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                 14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                 15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                 16:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                 17:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                 18:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1288	100.0	1288	15	US-10-006-780-2	Sequence 2, Appli
2	354	27.5	355	15	US-10-006-780-6	Sequence 6, Appli
3	332	25.8	332	15	US-10-006-780-4	Sequence 4, Appli
4	332	25.8	332	15	US-10-006-780-10	Sequence 10, Appl
5	330	25.6	361	15	US-10-006-780-8	Sequence 8, Appli
6	17	1.3	666	10	US-09-801-368-36	Sequence 36, Appl
7	16	1.2	213	12	US-10-032-585-7714	Sequence 7714, Ap
8	16	1.2	732	10	US-09-994-485-8	Sequence 8, Appli
9	16	1.2	732	10	US-09-832-292-12	Sequence 12, Appl
10	16	1.2	758	10	US-09-801-368-224	Sequence 224, App
11	16	1.2	798	12	US-10-032-585-7565	Sequence 7565, Ap
12	16	1.2	961	10	US-09-801-368-132	Sequence 132, App
13	16	1.2	1331	15	US-10-087-464-50	Sequence 50, Appl
14	16	1.2	2150	14	US-10-135-322-17	Sequence 17, Appl
15	16	1.2	2184	12	US-10-304-095-6	Sequence 6, Appli
16	16	1.2	2690	15	US-10-184-644-35	Sequence 35, Appl
17	16	1.2	2690	15	US-10-184-634-35	Sequence 35, Appl
18	15	1.2	73	9	US-09-864-761-47535	Sequence 47535, A
19	15	1.2	303	12	US-10-422-866-4	Sequence 4, Appli
20	15	1.2	492	12	US-10-422-866-2	Sequence 2, Appli
21	15	1.2	794	12	US-10-304-095-8	Sequence 8, Appli
22	15	1.2	888	9	US-09-826-752-6	Sequence 6, Appli
23	14	1.1	375	9	US-09-839-185-2	Sequence 2, Appli
24	14	1.1	458	12	US-10-032-585-7575	Sequence 7575, Ap
25	14	1.1	1062	15	US-10-195-144-37	Sequence 37, Appl
26	14	1.1	1770	10	US-09-801-368-298	Sequence 298, App
27	13	1.0	397	12	US-10-032-585-7237	Sequence 7237, Ap
28	13	1.0	941	12	US-10-032-585-7930	Sequence 7930, Ap
29	13	1.0	1151	10	US-09-801-368-134	Sequence 134, App
30	13	1.0	1151	15	US-10-060-019-31	Sequence 31, Appl
31	12	0.9	387	12	US-10-032-585-7034	Sequence 7034, Ap
32	12	0.9	410	10	US-09-925-300-1074	Sequence 1074, Ap
33	12	0.9	511	12	US-10-032-585-7788	Sequence 7788, Ap
34	12	0.9	725	12	US-09-849-602-20	Sequence 20, Appl
35	12	0.9	1501	10	US-09-924-154-17	Sequence 17, Appl
36	11	0.9	160	15	US-10-106-698-4938	Sequence 4938, Ap
37	11	0.9	250	11	US-09-934-455-462	Sequence 462, App
38	11	0.9	298	15	US-10-286-264-36	Sequence 36, Appl
39	11	0.9	298	16	US-10-278-536-188	Sequence 188, App
40	11	0.9	309	10	US-09-991-496-134	Sequence 134, App
41	11	0.9	314	11	US-09-893-519A-30	Sequence 30, Appl
42	11	0.9	319	10	US-09-723-276-2	Sequence 2, Appli
43	11	0.9	338	10	US-09-883-096-4	Sequence 4, Appli
44	11	0.9	343	15	US-10-159-151-8	Sequence 8, Appli
45	11	0.9	363	15	US-10-159-151-10	Sequence 10, Appl
46	11	0.9	373	15	US-10-159-151-4	Sequence 4, Appli
47	11	0.9	381	15	US-10-159-151-6	Sequence 6, Appli
48	11	0.9	383	10	US-09-883-096-5	Sequence 5, Appli
49	11	0.9	409	12	US-10-032-585-7054	Sequence 7054, Ap
50	11	0.9	676	10	US-09-801-368-302	Sequence 302, App
51	11	0.9	864	10	US-09-883-096-2	Sequence 2, Appli
52	11	0.9	1230	12	US-10-093-524-6	Sequence 6, Appli
53	11	0.9	1269	12	US-10-342-224-62	Sequence 62, Appl
54	11	0.9	1368	15	US-10-159-151-2	Sequence 2, Appli
55	11	0.9	1388	15	US-10-146-473-82	Sequence 82, Appl
56	10	0.8	16	11	US-09-539-382-40	Sequence 40, Appl
57	10	0.8	16	15	US-10-067-790-40	Sequence 40, Appl
58	10	0.8	16	15	US-10-067-892-40	Sequence 40, Appl
59	10	0.8	16	15	US-10-067-893-40	Sequence 40, Appl
60	10	0.8	40	12	US-10-083-768-226	Sequence 226, App
61	10	0.8	265	12	US-10-032-585-7166	Sequence 7166, Ap
62	10	0.8	412	10	US-09-788-345-12	Sequence 12, Appl
63	10	0.8	412	12	US-10-337-312-12	Sequence 12, Appl
64	10	0.8	488	12	US-10-032-585-7710	Sequence 7710, Ap
65	10	0.8	500	12	US-10-320-646-29	Sequence 29, Appl
66	10	0.8	500	12	US-10-320-646-34	Sequence 34, Appl
67	10	0.8	501	12	US-10-320-646-27	Sequence 27, Appl
68	10	0.8	501	12	US-10-320-646-28	Sequence 28, Appl
69	10	0.8	511	9	US-09-949-780-2	Sequence 2, Appli
70	10	0.8	522	9	US-09-965-313-7	Sequence 7, Appli
71	10	0.8	523	12	US-10-017-161-1982	Sequence 1982, Ap
72	10	0.8	544	12	US-10-032-585-7849	Sequence 7849, Ap
73	10	0.8	596	12	US-10-167-831-2	Sequence 2, Appli
74	10	0.8	623	14	US-10-108-605-125	Sequence 125, App
75	10	0.8	623	14	US-10-108-605-129	Sequence 129, App
76	10	0.8	659	10	US-09-833-555-20	Sequence 20, Appl
77	10	0.8	690	9	US-09-334-477-47	Sequence 47, Appl
78	10	0.8	694	9	US-09-334-477-49	Sequence 49, Appl
79	10	0.8	708	9	US-09-334-477-33	Sequence 33, Appl
80	10	0.8	711	9	US-09-334-477-35	Sequence 35, Appl
81	10	0.8	824	10	US-09-801-368-312	Sequence 312, App
82	10	0.8	861	11	US-09-820-843A-109	Sequence 109, App
83	10	0.8	1068	10	US-09-801-368-242	Sequence 242, App
84	10	0.8	1420	12	US-10-241-596-110	Sequence 110, App
85	10	0.8	1598	12	US-10-032-585-7766	Sequence 7766, Ap
86	10	0.8	1648	15	US-10-157-031-38	Sequence 38, Appl
87	9	0.7	34	9	US-09-864-761-48088	Sequence 48088, A
88	9	0.7	57	9	US-09-864-761-43371	Sequence 43371, A

89	9	0.7	177	12	US-09-903-190-164	Sequence 164, App
90	9	0.7	354	12	US-10-032-585-7079	Sequence 7079, Ap
91	9	0.7	440	15	US-10-083-357-1266	Sequence 1266, Ap
92	9	0.7	454	12	US-10-032-585-7816	Sequence 7816, Ap
93	9	0.7	498	12	US-10-032-585-7230	Sequence 7230, Ap
94	9	0.7	564	10	US-09-801-368-190	Sequence 190, App
95	9	0.7	670	12	US-10-032-585-7568	Sequence 7568, Ap
96	9	0.7	715	12	US-10-171-404A-44	Sequence 44, Appl
97	9	0.7	740	12	US-10-032-585-7673	Sequence 7673, Ap
98	9	0.7	764	12	US-10-032-585-7597	Sequence 7597, Ap
99	9	0.7	769	12	US-10-032-585-7117	Sequence 7117, Ap
100	9	0.7	807	11	US-09-820-843A-108	Sequence 108, App
101	9	0.7	856	12	US-09-949-029-66	Sequence 66, Appl
102	9	0.7	887	12	US-09-949-029-130	Sequence 130, App
103	9	0.7	928	10	US-09-801-368-282	Sequence 282, App
104	9	0.7	1143	12	US-10-032-585-7789	Sequence 7789, Ap
105	9	0.7	1165	14	US-10-016-768-10	Sequence 10, Appl
106	9	0.7	1207	12	US-09-949-029-144	Sequence 144, App
107	9	0.7	1221	14	US-10-016-768-11	Sequence 11, Appl
108	9	0.7	1245	12	US-10-032-585-7657	Sequence 7657, Ap
109	9	0.7	3092	10	US-09-801-368-172	Sequence 172, App
110	8	0.6	10	12	US-09-572-270A-565	Sequence 565, App
111	8	0.6	39	9	US-09-864-761-41774	Sequence 41774, A
112	8	0.6	41	9	US-09-864-761-44443	Sequence 44443, A
113	8	0.6	52	9	US-09-864-761-33973	Sequence 33973, A
114	8	0.6	75	9	US-09-864-761-44431	Sequence 44431, A
115	8	0.6	90	15	US-10-253-007-59	Sequence 59, Appl
116	8	0.6	146	15	US-10-180-375-138	Sequence 138, App
117	8	0.6	164	9	US-09-745-763-104	Sequence 104, App
118	8	0.6	175	14	US-10-108-605-97	Sequence 97, Appl
119	8	0.6	195	12	US-10-314-372-4	Sequence 4, Appli
120	8	0.6	221	15	US-10-021-811-14	Sequence 14, Appl
121	8	0.6	222	12	US-10-032-585-7349	Sequence 7349, Ap
122	8	0.6	229	9	US-09-925-301-1283	Sequence 1283, Ap
123	8	0.6	234	8	US-08-808-031A-35	Sequence 35, Appl
124	8	0.6	250	11	US-09-934-455-256	Sequence 256, App
125	8	0.6	250	15	US-10-180-375-194	Sequence 194, App
126	8	0.6	251	12	US-10-032-585-7420	Sequence 7420, Ap
127	8	0.6	251	15	US-10-295-403-142	Sequence 142, App
128	8	0.6	256	9	US-09-925-301-1083	Sequence 1083, Ap
129	8	0.6	259	8	US-08-808-031A-5	Sequence 5, Appli
130	8	0.6	300	9	US-09-861-451A-42	Sequence 42, Appl
131	8	0.6	302	9	US-09-443-704-12	Sequence 12, Appl
132	8	0.6	302	14	US-10-008-118A-12	Sequence 12, Appl
133	8	0.6	312	11	US-09-934-455-160	Sequence 160, App
134	8	0.6	324	9	US-09-874-923-53	Sequence 53, Appl
135	8	0.6	324	10	US-09-991-496-53	Sequence 53, Appl
136	8	0.6	333	15	US-10-278-173-146	Sequence 146, App
137	8	0.6	341	12	US-10-017-161-2038	Sequence 2038, Ap
138	8	0.6	342	15	US-10-132-652-3	Sequence 3, Appli
139	8	0.6	354	10	US-09-801-368-414	Sequence 414, App
140	8	0.6	372	15	US-10-132-652-2	Sequence 2, Appli
141	8	0.6	376	15	US-10-278-173-142	Sequence 142, App
142	8	0.6	408	9	US-09-802-213-9	Sequence 9, Appli
143	8	0.6	419	11	US-09-934-455-422	Sequence 422, App
144	8	0.6	420	12	US-10-032-585-7653	Sequence 7653, Ap
145	8	0.6	430	12	US-09-769-736-6	Sequence 6, Appli
146	8	0.6	464	12	US-10-161-051-119	Sequence 119, App
147	8	0.6	528	12	US-10-032-585-7571	Sequence 7571, Ap
148	8	0.6	554	10	US-09-801-368-142	Sequence 142, App
149	8	0.6	555	8	US-08-808-031A-31	Sequence 31, Appl
150	8	0.6	586	8	US-08-808-031A-28	Sequence 28, Appl
151	8	0.6	605	10	US-09-758-269-16	Sequence 16, Appl
152	8	0.6	652	9	US-09-351-794A-2	Sequence 2, Appli
153	8	0.6	653	9	US-09-746-801A-13	Sequence 13, Appl
154	8	0.6	710	12	US-10-032-585-7813	Sequence 7813, Ap
155	8	0.6	721	15	US-10-286-264-70	Sequence 70, Appl
156	8	0.6	725	11	US-09-934-455-442	Sequence 442, App
157	8	0.6	770	9	US-09-303-232-2	Sequence 2, Appli
158	8	0.6	809	12	US-10-032-585-7691	Sequence 7691, Ap
159	8	0.6	834	9	US-09-826-752-4	Sequence 4, Appli
160	8	0.6	854	12	US-10-335-711-9	Sequence 9, Appli
161	8	0.6	945	12	US-10-032-585-7114	Sequence 7114, Ap
162	8	0.6	1015	12	US-10-032-585-7162	Sequence 7162, Ap
163	8	0.6	1115	12	US-10-335-711-2	Sequence 2, Appli
164	8	0.6	1207	14	US-10-108-605-71	Sequence 71, Appl
165	8	0.6	1242	10	US-09-903-248-5	Sequence 5, Appli
166	8	0.6	1242	10	US-09-859-604-5	Sequence 5, Appli
167	8	0.6	1242	10	US-09-903-063-5	Sequence 5, Appli
168	8	0.6	1242	10	US-09-903-216-5	Sequence 5, Appli
169	8	0.6	1242	10	US-09-903-199-5	Sequence 5, Appli
170	8	0.6	1242	10	US-09-903-023-5	Sequence 5, Appli
171	8	0.6	1242	11	US-09-436-184-5	Sequence 5, Appli
172	8	0.6	1242	14	US-10-085-027-1	Sequence 1, Appli
173	8	0.6	1284	9	US-09-119-855-11	Sequence 11, Appl
174	8	0.6	1284	15	US-10-160-224-9	Sequence 9, Appli
175	8	0.6	1284	15	US-10-185-867-11	Sequence 11, Appl
176	8	0.6	1403	14	US-10-108-605-93	Sequence 93, Appl
177	8	0.6	1850	12	US-10-146-657-5	Sequence 5, Appli
178	7	0.5	10	12	US-09-572-270A-530	Sequence 530, App
179	7	0.5	10	12	US-09-572-270A-532	Sequence 532, App
180	7	0.5	10	12	US-09-572-270A-534	Sequence 534, App
181	7	0.5	10	12	US-09-572-270A-536	Sequence 536, App
182	7	0.5	10	12	US-09-572-270A-538	Sequence 538, App
183	7	0.5	10	12	US-09-572-270A-544	Sequence 544, App
184	7	0.5	10	12	US-09-572-270A-546	Sequence 546, App
185	7	0.5	10	12	US-09-572-270A-548	Sequence 548, App
186	7	0.5	10	12	US-09-572-270A-550	Sequence 550, App
187	7	0.5	10	12	US-09-572-270A-552	Sequence 552, App
188	7	0.5	10	12	US-09-572-270A-554	Sequence 554, App
189	7	0.5	10	12	US-09-572-270A-563	Sequence 563, App
190	7	0.5	10	12	US-09-572-270A-567	Sequence 567, App
191	7	0.5	10	12	US-09-572-270A-834	Sequence 834, App
192	7	0.5	10	12	US-09-572-270A-838	Sequence 838, App
193	7	0.5	10	12	US-09-572-270A-1008	Sequence 1008, Ap
194	7	0.5	10	12	US-09-572-270A-1010	Sequence 1010, Ap
195	7	0.5	16	9	US-09-935-682-11	Sequence 11, Appl
196	7	0.5	17	9	US-09-864-761-42401	Sequence 42401, A
197	7	0.5	17	11	US-09-876-904A-233	Sequence 233, App
198	7	0.5	19	9	US-09-864-761-37256	Sequence 37256, A
199	7	0.5	22	9	US-09-864-761-40447	Sequence 40447, A
200	7	0.5	23	11	US-09-967-869A-1	Sequence 1, Appli
201	7	0.5	24	14	US-10-123-427-7	Sequence 7, Appli
202	7	0.5	25	9	US-09-864-761-45219	Sequence 45219, A
203	7	0.5	26	9	US-09-864-761-44765	Sequence 44765, A
204	7	0.5	30	9	US-09-864-761-41595	Sequence 41595, A
205	7	0.5	30	9	US-09-864-761-42325	Sequence 42325, A
206	7	0.5	31	10	US-09-764-846-248	Sequence 248, App
207	7	0.5	31	15	US-10-091-483-248	Sequence 248, App
208	7	0.5	32	9	US-09-864-761-36029	Sequence 36029, A
209	7	0.5	33	9	US-09-864-761-36143	Sequence 36143, A
210	7	0.5	33	9	US-09-864-761-47462	Sequence 47462, A
211	7	0.5	43	14	US-10-124-557-12	Sequence 12, Appl
212	7	0.5	44	9	US-09-864-761-43901	Sequence 43901, A
213	7	0.5	46	9	US-09-864-761-45070	Sequence 45070, A
214	7	0.5	51	12	US-10-008-524A-100	Sequence 100, App
215	7	0.5	52	9	US-09-864-761-43647	Sequence 43647, A
216	7	0.5	54	9	US-09-864-761-34566	Sequence 34566, A
217	7	0.5	58	9	US-09-726-643-59	Sequence 59, Appl
218	7	0.5	58	14	US-10-042-141-59	Sequence 59, Appl
219	7	0.5	63	9	US-09-877-258-2	Sequence 2, Appli
220	7	0.5	63	9	US-09-864-761-42311	Sequence 42311, A
221	7	0.5	63	10	US-09-764-846-260	Sequence 260, App
222	7	0.5	63	15	US-10-091-483-260	Sequence 260, App
223	7	0.5	66	10	US-09-764-877-1302	Sequence 1302, Ap
224	7	0.5	71	10	US-09-738-626-6442	Sequence 6442, Ap
225	7	0.5	72	9	US-09-764-869-683	Sequence 683, App
226	7	0.5	72	15	US-10-091-504-683	Sequence 683, App
227	7	0.5	74	15	US-10-178-213-368	Sequence 368, App
228	7	0.5	94	10	US-09-764-864-1273	Sequence 1273, Ap
229	7	0.5	100	9	US-09-922-138-20	Sequence 20, Appl
230	7	0.5	101	9	US-09-858-664A-10	Sequence 10, Appl
231	7	0.5	104	9	US-09-864-761-41522	Sequence 41522, A
232	7	0.5	104	10	US-09-764-864-1276	Sequence 1276, Ap
233	7	0.5	105	14	US-10-135-322-6	Sequence 6, Appli
234	7	0.5	108	10	US-09-764-864-817	Sequence 817, App

235	7	0.5	109	9	US-09-764-853-625	Sequence 625, App
236	7	0.5	109	11	US-09-989-442-114	Sequence 114, App
237	7	0.5	109	15	US-10-073-865-101	Sequence 101, App
238	7	0.5	110	9	US-09-765-272-102	Sequence 102, App
239	7	0.5	111	9	US-09-858-664A-8	Sequence 8, Appli
240	7	0.5	112	10	US-09-764-864-812	Sequence 812, App
241	7	0.5	112	14	US-10-027-559-8	Sequence 8, Appli
242	7	0.5	118	15	US-10-106-698-4760	Sequence 4760, Ap
243	7	0.5	119	9	US-09-740-288A-10	Sequence 10, Appl
244	7	0.5	120	15	US-10-101-464A-641	Sequence 641, App
245	7	0.5	130	14	US-10-071-751-39	Sequence 39, Appl
246	7	0.5	132	14	US-10-078-929-36	Sequence 36, Appl
247	7	0.5	134	10	US-09-738-626-5042	Sequence 5042, Ap
248	7	0.5	139	11	US-09-934-455-6	Sequence 6, Appli
249	7	0.5	148	15	US-10-012-542-453	Sequence 453, App
250	7	0.5	150	11	US-09-934-455-86	Sequence 86, Appl
251	7	0.5	152	11	US-09-974-879-542	Sequence 542, App
252	7	0.5	152	11	US-09-305-736-544	Sequence 544, App
253	7	0.5	156	9	US-09-925-301-1154	Sequence 1154, Ap
254	7	0.5	165	10	US-09-764-864-1267	Sequence 1267, Ap
255	7	0.5	166	12	US-10-076-747-83	Sequence 83, Appl
256	7	0.5	169	9	US-09-864-761-42972	Sequence 42972, A
257	7	0.5	172	14	US-10-124-557-88	Sequence 88, Appl
258	7	0.5	173	10	US-09-738-626-6015	Sequence 6015, Ap
259	7	0.5	176	15	US-10-106-698-5906	Sequence 5906, Ap
260	7	0.5	180	9	US-09-910-150-29	Sequence 29, Appl
261	7	0.5	181	14	US-10-062-254-369	Sequence 369, App
262	7	0.5	181	15	US-10-149-819-14	Sequence 14, Appl
263	7	0.5	185	12	US-10-342-224-80	Sequence 80, Appl
264	7	0.5	187	12	US-09-903-190-102	Sequence 102, App
265	7	0.5	191	11	US-09-934-455-54	Sequence 54, Appl
266	7	0.5	192	10	US-09-764-868-1192	Sequence 1192, Ap
267	7	0.5	192	14	US-10-124-557-90	Sequence 90, Appl
268	7	0.5	194	9	US-09-910-150-28	Sequence 28, Appl
269	7	0.5	195	10	US-09-764-864-854	Sequence 854, App
270	7	0.5	199	10	US-09-764-864-1291	Sequence 1291, Ap
271	7	0.5	200	9	US-09-867-550-1934	Sequence 1934, Ap
272	7	0.5	201	10	US-09-764-864-836	Sequence 836, App
273	7	0.5	204	11	US-09-934-455-84	Sequence 84, Appl
274	7	0.5	204	14	US-10-124-557-92	Sequence 92, Appl
275	7	0.5	207	10	US-09-965-602-34	Sequence 34, Appl
276	7	0.5	207	11	US-09-975-719-211	Sequence 211, App
277	7	0.5	207	14	US-10-062-254-174	Sequence 174, App
278	7	0.5	208	9	US-09-733-507-13	Sequence 13, Appl
279	7	0.5	208	14	US-10-124-557-132	Sequence 132, App
280	7	0.5	209	14	US-10-124-557-94	Sequence 94, Appl
281	7	0.5	211	10	US-09-908-805B-10	Sequence 10, Appl
282	7	0.5	211	12	US-10-331-695-2	Sequence 2, Appli
283	7	0.5	212	9	US-09-350-874-63	Sequence 63, Appl
284	7	0.5	212	10	US-09-908-805B-4	Sequence 4, Appli
285	7	0.5	212	10	US-09-908-805B-12	Sequence 12, Appl
286	7	0.5	212	12	US-10-331-695-4	Sequence 4, Appli
287	7	0.5	212	15	US-10-038-010-60	Sequence 60, Appl
288	7	0.5	212	15	US-10-278-173-122	Sequence 122, App
289	7	0.5	212	15	US-10-295-403-16	Sequence 16, Appl
290	7	0.5	212	15	US-10-106-989-63	Sequence 63, Appl
291	7	0.5	214	9	US-09-925-302-649	Sequence 649, App
292	7	0.5	214	12	US-10-293-971-16	Sequence 16, Appl
293	7	0.5	216	12	US-10-267-430-2	Sequence 2, Appli
294	7	0.5	217	14	US-10-124-557-76	Sequence 76, Appl
295	7	0.5	217	15	US-10-021-811-63	Sequence 63, Appl
296	7	0.5	217	15	US-10-106-698-5191	Sequence 5191, Ap
297	7	0.5	217	15	US-10-106-698-5795	Sequence 5795, Ap
298	7	0.5	218	11	US-09-934-455-192	Sequence 192, App
299	7	0.5	218	13	US-10-010-901-31	Sequence 31, Appl
300	7	0.5	220	14	US-10-124-557-96	Sequence 96, Appl
301	7	0.5	220	15	US-10-101-464A-628	Sequence 628, App
302	7	0.5	221	12	US-10-032-585-7567	Sequence 7567, Ap
303	7	0.5	222	15	US-10-156-761-7686	Sequence 7686, Ap
304	7	0.5	225	15	US-10-156-761-12311	Sequence 12311, A
305	7	0.5	226	11	US-09-967-869A-15	Sequence 15, Appl
306	7	0.5	229	12	US-10-411-224-130	Sequence 130, App
307	7	0.5	229	15	US-10-091-333-9	Sequence 9, Appli
308	7	0.5	229	15	US-10-325-878-9	Sequence 9, Appli
309	7	0.5	231	14	US-10-124-557-30	Sequence 30, Appl
310	7	0.5	238	15	US-10-195-117-3	Sequence 3, Appli
311	7	0.5	238	15	US-10-195-117-4	Sequence 4, Appli
312	7	0.5	242	12	US-10-298-638-17	Sequence 17, Appl
313	7	0.5	243	11	US-09-533-029-34	Sequence 34, Appl
314	7	0.5	243	11	US-09-934-455-416	Sequence 416, App
315	7	0.5	246	9	US-09-864-761-36100	Sequence 36100, A
316	7	0.5	248	15	US-10-103-313-442	Sequence 442, App
317	7	0.5	256	15	US-10-156-761-11972	Sequence 11972, A
318	7	0.5	259	11	US-09-934-455-298	Sequence 298, App
319	7	0.5	260	11	US-09-906-393A-18	Sequence 18, Appl
320	7	0.5	261	12	US-10-032-201B-162	Sequence 162, App
321	7	0.5	261	14	US-10-062-254-140	Sequence 140, App
322	7	0.5	261	14	US-10-062-254-142	Sequence 142, App
323	7	0.5	262	11	US-09-906-393A-14	Sequence 14, Appl
324	7	0.5	262	11	US-09-906-393A-16	Sequence 16, Appl
325	7	0.5	270	14	US-10-078-929-38	Sequence 38, Appl
326	7	0.5	270	15	US-10-286-264-96	Sequence 96, Appl
327	7	0.5	273	9	US-09-815-242-5550	Sequence 5550, Ap
328	7	0.5	282	15	US-10-295-403-54	Sequence 54, Appl
329	7	0.5	284	11	US-09-056-019-3	Sequence 3, Appli
330	7	0.5	284	11	US-09-934-455-104	Sequence 104, App
331	7	0.5	285	15	US-10-157-031-132	Sequence 132, App
332	7	0.5	286	9	US-09-815-242-12234	Sequence 12234, A
333	7	0.5	288	12	US-09-896-186B-24	Sequence 24, Appl
334	7	0.5	290	15	US-10-106-698-5388	Sequence 5388, Ap
335	7	0.5	295	16	US-10-278-536-184	Sequence 184, App
336	7	0.5	296	14	US-10-124-557-70	Sequence 70, Appl
337	7	0.5	297	16	US-10-278-536-74	Sequence 74, Appl
338	7	0.5	302	11	US-09-725-311-2	Sequence 2, Appli
339	7	0.5	306	12	US-10-166-225A-101	Sequence 101, App
340	7	0.5	307	12	US-10-238-075-453	Sequence 453, App
341	7	0.5	307	15	US-10-278-173-148	Sequence 148, App
342	7	0.5	307	16	US-10-278-536-92	Sequence 92, Appl
343	7	0.5	310	15	US-10-007-521-22	Sequence 22, Appl
344	7	0.5	311	12	US-10-259-165-318	Sequence 318, App
345	7	0.5	312	10	US-09-738-626-5731	Sequence 5731, Ap
346	7	0.5	313	12	US-09-896-186B-2	Sequence 2, Appli
347	7	0.5	314	9	US-09-768-826-43	Sequence 43, Appl
348	7	0.5	315	15	US-10-232-563-11	Sequence 11, Appl
349	7	0.5	319	12	US-10-168-080-2	Sequence 2, Appli
350	7	0.5	321	11	US-09-934-455-418	Sequence 418, App
351	7	0.5	323	12	US-10-017-161-1620	Sequence 1620, Ap
352	7	0.5	324	9	US-09-734-417-2	Sequence 2, Appli
353	7	0.5	325	11	US-09-934-455-138	Sequence 138, App
354	7	0.5	328	10	US-09-900-527-2	Sequence 2, Appli
355	7	0.5	328	15	US-10-157-031-12	Sequence 12, Appl
356	7	0.5	330	11	US-09-725-311-4	Sequence 4, Appli
357	7	0.5	339	11	US-09-967-869A-14	Sequence 14, Appl
358	7	0.5	343	16	US-10-212-677-247	Sequence 247, App
359	7	0.5	343	16	US-10-212-677-248	Sequence 248, App
360	7	0.5	343	16	US-10-212-677-249	Sequence 249, App
361	7	0.5	347	10	US-09-816-127-3	Sequence 3, Appli
362	7	0.5	349	9	US-09-802-213-7	Sequence 7, Appli
363	7	0.5	349	11	US-09-820-598-3	Sequence 3, Appli
364	7	0.5	349	14	US-10-095-932-3	Sequence 3, Appli
365	7	0.5	351	9	US-09-350-874-61	Sequence 61, Appl
366	7	0.5	351	15	US-10-106-989-61	Sequence 61, Appl
367	7	0.5	352	12	US-10-259-165-216	Sequence 216, App
368	7	0.5	360	9	US-09-817-647-7	Sequence 7, Appli
369	7	0.5	360	10	US-09-877-665-7	Sequence 7, Appli
370	7	0.5	360	14	US-10-136-573A-7	Sequence 7, Appli
371	7	0.5	360	15	US-10-215-862-7	Sequence 7, Appli
372	7	0.5	368	9	US-09-805-550-4	Sequence 4, Appli
373	7	0.5	368	10	US-09-801-368-308	Sequence 308, App
374	7	0.5	369	15	US-10-036-542-84	Sequence 84, Appl
375	7	0.5	372	14	US-10-124-557-64	Sequence 64, Appl
376	7	0.5	373	11	US-09-819-142-2	Sequence 2, Appli
377	7	0.5	373	11	US-09-934-455-108	Sequence 108, App
378	7	0.5	379	9	US-09-740-288A-26	Sequence 26, Appl
379	7	0.5	383	12	US-10-032-201B-196	Sequence 196, App
380	7	0.5	383	15	US-10-141-531-72	Sequence 72, Appl

381	7	0.5	383	15	US-10-141-531-87	Sequence 87, Appl
382	7	0.5	383	15	US-10-141-531-96	Sequence 96, Appl
383	7	0.5	383	15	US-10-141-531-105	Sequence 105, App
384	7	0.5	383	15	US-10-141-531-114	Sequence 114, App
385	7	0.5	383	15	US-10-141-531-123	Sequence 123, App
386	7	0.5	383	15	US-10-141-531-132	Sequence 132, App
387	7	0.5	383	15	US-10-141-531-141	Sequence 141, App
388	7	0.5	383	15	US-10-141-531-150	Sequence 150, App
389	7	0.5	383	15	US-10-141-531-159	Sequence 159, App
390	7	0.5	383	15	US-10-141-531-168	Sequence 168, App
391	7	0.5	383	15	US-10-141-531-177	Sequence 177, App
392	7	0.5	383	15	US-10-141-531-186	Sequence 186, App
393	7	0.5	383	15	US-10-141-531-195	Sequence 195, App
394	7	0.5	383	15	US-10-141-531-204	Sequence 204, App
395	7	0.5	384	10	US-09-945-249-8	Sequence 8, Appli
396	7	0.5	385	12	US-10-183-708-96	Sequence 96, Appl
397	7	0.5	387	12	US-10-082-272-3	Sequence 3, Appli
398	7	0.5	391	15	US-10-156-275-6	Sequence 6, Appli
399	7	0.5	392	10	US-09-168-474-9	Sequence 9, Appli
400	7	0.5	392	12	US-10-054-399A-9	Sequence 9, Appli
401	7	0.5	392	15	US-10-060-036-67	Sequence 67, Appl
402	7	0.5	395	11	US-09-949-427-2	Sequence 2, Appli
403	7	0.5	397	14	US-10-029-180-40	Sequence 40, Appl
404	7	0.5	398	10	US-09-961-527A-9	Sequence 9, Appli
405	7	0.5	399	11	US-09-934-455-278	Sequence 278, App
406	7	0.5	401	12	US-09-949-029-146	Sequence 146, App
407	7	0.5	403	9	US-09-802-213-5	Sequence 5, Appli
408	7	0.5	406	11	US-09-056-019-1	Sequence 1, Appli
409	7	0.5	411	15	US-10-156-761-11579	Sequence 11579, A
410	7	0.5	414	9	US-09-802-213-3	Sequence 3, Appli
411	7	0.5	414	9	US-09-802-213-10	Sequence 10, Appl
412	7	0.5	415	10	US-09-931-457A-64	Sequence 64, Appl
413	7	0.5	422	14	US-10-124-557-68	Sequence 68, Appl
414	7	0.5	423	14	US-10-124-557-66	Sequence 66, Appl
415	7	0.5	428	11	US-09-056-019-24	Sequence 24, Appl
416	7	0.5	431	9	US-09-815-242-5406	Sequence 5406, Ap
417	7	0.5	431	9	US-09-815-242-12488	Sequence 12488, A
418	7	0.5	433	10	US-09-801-368-38	Sequence 38, Appl
419	7	0.5	435	10	US-09-866-582-33	Sequence 33, Appl
420	7	0.5	438	10	US-09-738-626-6703	Sequence 6703, Ap
421	7	0.5	446	12	US-10-254-995-6	Sequence 6, Appli
422	7	0.5	446	12	US-10-254-995-9	Sequence 9, Appli
423	7	0.5	447	12	US-10-082-272-2	Sequence 2, Appli
424	7	0.5	448	11	US-09-934-455-126	Sequence 126, App
425	7	0.5	450	9	US-09-826-752-2	Sequence 2, Appli
426	7	0.5	450	15	US-10-153-668-268	Sequence 268, App
427	7	0.5	450	15	US-10-153-668-380	Sequence 380, App
428	7	0.5	450	15	US-10-153-668-382	Sequence 382, App
429	7	0.5	452	9	US-09-815-242-10493	Sequence 10493, A
430	7	0.5	455	12	US-09-848-841-12	Sequence 12, Appl
431	7	0.5	459	11	US-09-056-019-39	Sequence 39, Appl
432	7	0.5	460	10	US-09-925-300-1228	Sequence 1228, Ap
433	7	0.5	460	11	US-09-056-019-38	Sequence 38, Appl
434	7	0.5	463	14	US-10-124-557-54	Sequence 54, Appl
435	7	0.5	464	11	US-09-770-509-22	Sequence 22, Appl
436	7	0.5	465	9	US-09-884-260A-52	Sequence 52, Appl
437	7	0.5	466	12	US-10-032-585-7466	Sequence 7466, Ap
438	7	0.5	471	8	US-08-954-771-34	Sequence 34, Appl
439	7	0.5	471	9	US-09-741-669-428	Sequence 428, App
440	7	0.5	474	10	US-09-736-457-1812	Sequence 1812, Ap
441	7	0.5	474	10	US-09-902-941-1812	Sequence 1812, Ap
442	7	0.5	474	10	US-09-849-626-1812	Sequence 1812, Ap
443	7	0.5	474	12	US-10-113-872-1812	Sequence 1812, Ap
444	7	0.5	474	15	US-10-017-754-1812	Sequence 1812, Ap
445	7	0.5	474	15	US-10-205-823-381	Sequence 381, App
446	7	0.5	475	9	US-09-826-752-14	Sequence 14, Appl
447	7	0.5	475	15	US-10-318-780-4	Sequence 4, Appli
448	7	0.5	476	10	US-09-774-414-3	Sequence 3, Appli
449	7	0.5	477	12	US-10-161-051-43	Sequence 43, Appl
450	7	0.5	484	12	US-10-347-278-17	Sequence 17, Appl
451	7	0.5	484	12	US-10-347-252-17	Sequence 17, Appl
452	7	0.5	489	9	US-09-956-425-6	Sequence 6, Appli
453	7	0.5	489	11	US-09-533-029-60	Sequence 60, Appl
454	7	0.5	490	15	US-10-121-746-6	Sequence 6, Appli
455	7	0.5	491	11	US-09-999-220B-3	Sequence 3, Appli
456	7	0.5	491	11	US-09-999-220B-4	Sequence 4, Appli
457	7	0.5	491	12	US-10-032-585-7161	Sequence 7161, Ap
458	7	0.5	495	12	US-10-293-086-73	Sequence 73, Appl
459	7	0.5	500	15	US-10-006-852-16	Sequence 16, Appl
460	7	0.5	500	15	US-10-219-220-149	Sequence 149, App
461	7	0.5	500	15	US-10-228-897-13	Sequence 13, Appl
462	7	0.5	503	10	US-09-801-368-424	Sequence 424, App
463	7	0.5	506	10	US-09-976-165-19	Sequence 19, Appl
464	7	0.5	506	12	US-10-342-276-19	Sequence 19, Appl
465	7	0.5	511	11	US-09-298-523B-13	Sequence 13, Appl
466	7	0.5	511	15	US-10-128-714-3150	Sequence 3150, Ap
467	7	0.5	511	15	US-10-128-714-8150	Sequence 8150, Ap
468	7	0.5	511	15	US-10-156-761-9625	Sequence 9625, Ap
469	7	0.5	513	11	US-09-298-523B-12	Sequence 12, Appl
470	7	0.5	518	9	US-09-925-299-862	Sequence 862, App
471	7	0.5	518	11	US-09-925-299-862	Sequence 862, App
472	7	0.5	518	11	US-09-815-379-4	Sequence 4, Appli
473	7	0.5	518	15	US-10-283-023-2	Sequence 2, Appli
474	7	0.5	519	16	US-10-278-536-182	Sequence 182, App
475	7	0.5	522	15	US-10-081-051-107	Sequence 107, App
476	7	0.5	524	15	US-10-253-007-56	Sequence 56, Appl
477	7	0.5	530	12	US-10-032-585-7699	Sequence 7699, Ap
478	7	0.5	542	9	US-09-925-301-1004	Sequence 1004, Ap
479	7	0.5	543	10	US-09-736-457-337	Sequence 337, App
480	7	0.5	543	10	US-09-902-941-337	Sequence 337, App
481	7	0.5	543	10	US-09-849-626-337	Sequence 337, App
482	7	0.5	543	11	US-09-476-300-337	Sequence 337, App
483	7	0.5	543	12	US-10-251-661-6	Sequence 6, Appli
484	7	0.5	543	12	US-10-113-872-337	Sequence 337, App
485	7	0.5	543	12	US-10-032-585-7577	Sequence 7577, Ap
486	7	0.5	543	15	US-10-017-754-337	Sequence 337, App
487	7	0.5	543	15	US-10-205-823-95	Sequence 95, Appl
488	7	0.5	544	12	US-10-032-585-7924	Sequence 7924, Ap
489	7	0.5	553	12	US-10-347-278-13	Sequence 13, Appl
490	7	0.5	553	12	US-10-347-252-13	Sequence 13, Appl
491	7	0.5	555	15	US-10-106-698-5847	Sequence 5847, Ap
492	7	0.5	559	12	US-10-032-585-7914	Sequence 7914, Ap
493	7	0.5	562	15	US-10-083-357-1329	Sequence 1329, Ap
494	7	0.5	578	12	US-10-032-585-7793	Sequence 7793, Ap
495	7	0.5	579	12	US-10-384-743-4	Sequence 4, Appli
496	7	0.5	579	14	US-10-158-895-4	Sequence 4, Appli
497	7	0.5	582	15	US-10-122-805-3	Sequence 3, Appli
498	7	0.5	590	12	US-10-384-743-15	Sequence 15, Appl
499	7	0.5	590	14	US-10-158-895-15	Sequence 15, Appl
500	7	0.5	594	12	US-10-137-870-10	Sequence 10, Appl
501	7	0.5	594	12	US-10-140-018-10	Sequence 10, Appl
502	7	0.5	594	12	US-10-140-021-10	Sequence 10, Appl
503	7	0.5	594	12	US-10-140-274-10	Sequence 10, Appl
504	7	0.5	594	12	US-10-140-471-10	Sequence 10, Appl
505	7	0.5	594	12	US-10-140-807-10	Sequence 10, Appl
506	7	0.5	594	12	US-10-140-922-10	Sequence 10, Appl
507	7	0.5	594	12	US-10-140-924-10	Sequence 10, Appl
508	7	0.5	594	12	US-10-140-926-10	Sequence 10, Appl
509	7	0.5	594	12	US-10-141-698-10	Sequence 10, Appl
510	7	0.5	594	12	US-10-141-702-10	Sequence 10, Appl
511	7	0.5	594	12	US-10-141-704-10	Sequence 10, Appl
512	7	0.5	594	12	US-10-142-421-10	Sequence 10, Appl
513	7	0.5	594	12	US-10-142-432-10	Sequence 10, Appl
514	7	0.5	594	12	US-10-142-767-10	Sequence 10, Appl
515	7	0.5	594	12	US-10-143-033-10	Sequence 10, Appl
516	7	0.5	594	12	US-10-144-994-10	Sequence 10, Appl
517	7	0.5	594	12	US-10-145-628-10	Sequence 10, Appl
518	7	0.5	594	12	US-10-145-631-10	Sequence 10, Appl
519	7	0.5	594	12	US-10-145-633-10	Sequence 10, Appl
520	7	0.5	594	12	US-10-145-746-10	Sequence 10, Appl
521	7	0.5	594	12	US-10-145-748-10	Sequence 10, Appl
522	7	0.5	594	12	US-10-145-823-10	Sequence 10, Appl
523	7	0.5	594	12	US-10-145-826-10	Sequence 10, Appl
524	7	0.5	594	12	US-10-145-870-10	Sequence 10, Appl
525	7	0.5	594	12	US-10-145-876-10	Sequence 10, Appl
526	7	0.5	594	12	US-10-145-959-10	Sequence 10, Appl

527	7	0.5	594	12	US-10-146-724-10	Sequence 10, Appl
528	7	0.5	594	12	US-10-146-725-10	Sequence 10, Appl
529	7	0.5	594	12	US-10-146-795-10	Sequence 10, Appl
530	7	0.5	594	12	US-10-147-495-10	Sequence 10, Appl
531	7	0.5	594	12	US-10-147-501-10	Sequence 10, Appl
532	7	0.5	594	12	US-10-147-504-10	Sequence 10, Appl
533	7	0.5	594	12	US-10-147-506-10	Sequence 10, Appl
534	7	0.5	594	12	US-10-147-509-10	Sequence 10, Appl
535	7	0.5	594	12	US-10-147-510-10	Sequence 10, Appl
536	7	0.5	594	12	US-10-147-511-10	Sequence 10, Appl
537	7	0.5	594	12	US-10-147-529-10	Sequence 10, Appl
538	7	0.5	594	12	US-10-152-397-10	Sequence 10, Appl
539	7	0.5	594	12	US-10-153-586-10	Sequence 10, Appl
540	7	0.5	594	12	US-10-158-783-10	Sequence 10, Appl
541	7	0.5	594	12	US-10-158-786-10	Sequence 10, Appl
542	7	0.5	594	12	US-10-140-019-10	Sequence 10, Appl
543	7	0.5	594	12	US-10-140-022-10	Sequence 10, Appl
544	7	0.5	594	12	US-10-140-861-10	Sequence 10, Appl
545	7	0.5	594	12	US-10-140-862-10	Sequence 10, Appl
546	7	0.5	594	12	US-10-141-697-10	Sequence 10, Appl
547	7	0.5	594	12	US-10-141-700-10	Sequence 10, Appl
548	7	0.5	594	12	US-10-141-705-10	Sequence 10, Appl
549	7	0.5	594	12	US-10-141-753-10	Sequence 10, Appl
550	7	0.5	594	12	US-10-141-758-10	Sequence 10, Appl
551	7	0.5	594	12	US-10-142-418-10	Sequence 10, Appl
552	7	0.5	594	12	US-10-142-420-10	Sequence 10, Appl
553	7	0.5	594	12	US-10-142-422-10	Sequence 10, Appl
554	7	0.5	594	12	US-10-142-427-10	Sequence 10, Appl
555	7	0.5	594	12	US-10-142-760-10	Sequence 10, Appl
556	7	0.5	594	12	US-10-145-821-10	Sequence 10, Appl
557	7	0.5	594	12	US-10-152-531-10	Sequence 10, Appl
558	7	0.5	594	12	US-10-127-840A-10	Sequence 10, Appl
559	7	0.5	594	12	US-10-142-424-10	Sequence 10, Appl
560	7	0.5	594	12	US-10-142-761-10	Sequence 10, Appl
561	7	0.5	594	12	US-10-142-763-10	Sequence 10, Appl
562	7	0.5	594	12	US-10-142-765-10	Sequence 10, Appl
563	7	0.5	594	12	US-10-142-887-10	Sequence 10, Appl
564	7	0.5	594	12	US-10-142-888-10	Sequence 10, Appl
565	7	0.5	594	12	US-10-143-034-10	Sequence 10, Appl
566	7	0.5	594	12	US-10-143-116-10	Sequence 10, Appl
567	7	0.5	594	12	US-10-144-957-10	Sequence 10, Appl
568	7	0.5	594	12	US-10-144-992-10	Sequence 10, Appl
569	7	0.5	594	12	US-10-145-015-10	Sequence 10, Appl
570	7	0.5	594	12	US-10-145-090-10	Sequence 10, Appl
571	7	0.5	594	12	US-10-145-091-10	Sequence 10, Appl
572	7	0.5	594	12	US-10-145-629-10	Sequence 10, Appl
573	7	0.5	594	12	US-10-145-630-10	Sequence 10, Appl
574	7	0.5	594	12	US-10-145-747-10	Sequence 10, Appl
575	7	0.5	594	12	US-10-145-752-10	Sequence 10, Appl
576	7	0.5	594	12	US-10-145-754-10	Sequence 10, Appl
577	7	0.5	594	12	US-10-145-755-10	Sequence 10, Appl
578	7	0.5	594	12	US-10-145-818-10	Sequence 10, Appl
579	7	0.5	594	12	US-10-145-820-10	Sequence 10, Appl
580	7	0.5	594	12	US-10-145-872-10	Sequence 10, Appl
581	7	0.5	594	12	US-10-145-873-10	Sequence 10, Appl
582	7	0.5	594	12	US-10-147-481-10	Sequence 10, Appl
583	7	0.5	594	12	US-10-147-482-10	Sequence 10, Appl
584	7	0.5	594	12	US-10-147-503-10	Sequence 10, Appl
585	7	0.5	594	12	US-10-147-522-10	Sequence 10, Appl
586	7	0.5	594	12	US-10-152-401-10	Sequence 10, Appl
587	7	0.5	594	12	US-10-157-783-10	Sequence 10, Appl
588	7	0.5	594	12	US-10-158-462-10	Sequence 10, Appl
589	7	0.5	594	12	US-10-158-792-10	Sequence 10, Appl
590	7	0.5	594	12	US-10-143-035-10	Sequence 10, Appl
591	7	0.5	594	12	US-10-145-751-10	Sequence 10, Appl
592	7	0.5	594	12	US-10-145-822-10	Sequence 10, Appl
593	7	0.5	594	12	US-10-145-824-10	Sequence 10, Appl
594	7	0.5	594	12	US-10-145-827-10	Sequence 10, Appl
595	7	0.5	594	12	US-10-145-869-10	Sequence 10, Appl
596	7	0.5	594	12	US-10-145-875-10	Sequence 10, Appl
597	7	0.5	594	12	US-10-145-877-10	Sequence 10, Appl
598	7	0.5	594	12	US-10-145-958-10	Sequence 10, Appl
599	7	0.5	594	12	US-10-146-787-10	Sequence 10, Appl
600	7	0.5	594	12	US-10-146-790-10	Sequence 10, Appl
601	7	0.5	594	12	US-10-146-793-10	Sequence 10, Appl
602	7	0.5	594	12	US-10-147-480-10	Sequence 10, Appl
603	7	0.5	594	12	US-10-147-485-10	Sequence 10, Appl
604	7	0.5	594	12	US-10-147-486-10	Sequence 10, Appl
605	7	0.5	594	12	US-10-147-487-10	Sequence 10, Appl
606	7	0.5	594	12	US-10-147-490-10	Sequence 10, Appl
607	7	0.5	594	12	US-10-147-494-10	Sequence 10, Appl
608	7	0.5	594	12	US-10-147-498-10	Sequence 10, Appl
609	7	0.5	594	12	US-10-147-514-10	Sequence 10, Appl
610	7	0.5	594	12	US-10-147-524-10	Sequence 10, Appl
611	7	0.5	594	12	US-10-152-379-10	Sequence 10, Appl
612	7	0.5	594	12	US-10-152-394-10	Sequence 10, Appl
613	7	0.5	594	12	US-10-152-406-10	Sequence 10, Appl
614	7	0.5	594	12	US-10-156-847-10	Sequence 10, Appl
615	7	0.5	594	12	US-10-157-778-10	Sequence 10, Appl
616	7	0.5	594	12	US-10-157-799-10	Sequence 10, Appl
617	7	0.5	594	12	US-10-160-504-10	Sequence 10, Appl
618	7	0.5	594	12	US-10-145-634-10	Sequence 10, Appl
619	7	0.5	594	12	US-10-147-520-10	Sequence 10, Appl
620	7	0.5	594	12	US-10-157-781-10	Sequence 10, Appl
621	7	0.5	594	12	US-10-176-989-10	Sequence 10, Appl
622	7	0.5	594	12	US-10-147-491-10	Sequence 10, Appl
623	7	0.5	594	12	US-10-152-378-10	Sequence 10, Appl
624	7	0.5	594	12	US-10-152-382-10	Sequence 10, Appl
625	7	0.5	594	12	US-10-152-383-10	Sequence 10, Appl
626	7	0.5	594	12	US-10-152-384-10	Sequence 10, Appl
627	7	0.5	594	12	US-10-152-387-10	Sequence 10, Appl
628	7	0.5	594	12	US-10-152-389-10	Sequence 10, Appl
629	7	0.5	594	12	US-10-152-390-10	Sequence 10, Appl
630	7	0.5	594	12	US-10-152-392-10	Sequence 10, Appl
631	7	0.5	594	12	US-10-153-756-10	Sequence 10, Appl
632	7	0.5	594	12	US-10-157-779-10	Sequence 10, Appl
633	7	0.5	594	12	US-10-157-784-10	Sequence 10, Appl
634	7	0.5	594	12	US-10-157-797-10	Sequence 10, Appl
635	7	0.5	594	12	US-10-158-491-10	Sequence 10, Appl
636	7	0.5	594	12	US-10-139-980-10	Sequence 10, Appl
637	7	0.5	594	12	US-10-142-762-10	Sequence 10, Appl
638	7	0.5	594	12	US-10-142-764-10	Sequence 10, Appl
639	7	0.5	594	12	US-10-142-766-10	Sequence 10, Appl
640	7	0.5	594	12	US-10-145-625-10	Sequence 10, Appl
641	7	0.5	594	12	US-10-145-627-10	Sequence 10, Appl
642	7	0.5	594	12	US-10-145-960-10	Sequence 10, Appl
643	7	0.5	594	12	US-10-145-962-10	Sequence 10, Appl
644	7	0.5	594	12	US-10-146-789-10	Sequence 10, Appl
645	7	0.5	594	12	US-10-147-483-10	Sequence 10, Appl
646	7	0.5	594	12	US-10-147-496-10	Sequence 10, Appl
647	7	0.5	594	12	US-10-147-505-10	Sequence 10, Appl
648	7	0.5	594	12	US-10-147-516-10	Sequence 10, Appl
649	7	0.5	594	12	US-10-152-398-10	Sequence 10, Appl
650	7	0.5	594	15	US-10-028-072-10	Sequence 10, Appl
651	7	0.5	594	15	US-10-121-049-10	Sequence 10, Appl
652	7	0.5	594	15	US-10-123-904-10	Sequence 10, Appl
653	7	0.5	594	15	US-10-140-470-10	Sequence 10, Appl
654	7	0.5	594	15	US-10-175-746-10	Sequence 10, Appl
655	7	0.5	594	15	US-10-176-918-10	Sequence 10, Appl
656	7	0.5	594	15	US-10-176-921-10	Sequence 10, Appl
657	7	0.5	594	15	US-10-137-865-10	Sequence 10, Appl
658	7	0.5	594	15	US-10-140-474-10	Sequence 10, Appl
659	7	0.5	594	15	US-10-142-431-10	Sequence 10, Appl
660	7	0.5	594	15	US-10-143-114-10	Sequence 10, Appl
661	7	0.5	594	15	US-10-140-002-10	Sequence 10, Appl
662	7	0.5	594	15	US-10-142-419-10	Sequence 10, Appl
663	7	0.5	594	15	US-10-123-262-10	Sequence 10, Appl
664	7	0.5	594	15	US-10-142-423-10	Sequence 10, Appl
665	7	0.5	594	15	US-10-121-050-10	Sequence 10, Appl
666	7	0.5	594	15	US-10-141-755-10	Sequence 10, Appl
667	7	0.5	594	15	US-10-143-032-10	Sequence 10, Appl
668	7	0.5	594	15	US-10-123-108-10	Sequence 10, Appl
669	7	0.5	594	15	US-10-123-236-10	Sequence 10, Appl
670	7	0.5	594	15	US-10-123-261-10	Sequence 10, Appl
671	7	0.5	594	15	US-10-140-921-10	Sequence 10, Appl
672	7	0.5	594	15	US-10-140-928-10	Sequence 10, Appl

673	7	0.5	594	15	US-10-121-045-10 Sequence 10, Appl
674	7	0.5	594	15	US-10-123-292-10 Sequence 10, Appl
675	7	0.5	594	15	US-10-123-903-10 Sequence 10, Appl
676	7	0.5	594	15	US-10-124-819-10 Sequence 10, Appl
677	7	0.5	594	15	US-10-124-822-10 Sequence 10, Appl
678	7	0.5	594	15	US-10-140-925-10 Sequence 10, Appl
679	7	0.5	594	15	US-10-160-498-10 Sequence 10, Appl
680	7	0.5	594	15	US-10-124-824-10 Sequence 10, Appl
681	7	0.5	594	15	US-10-127-825A-10 Sequence 10, Appl
682	7	0.5	594	15	US-10-127-829A-10 Sequence 10, Appl
683	7	0.5	594	15	US-10-127-835A-10 Sequence 10, Appl
684	7	0.5	594	15	US-10-127-839A-10 Sequence 10, Appl
685	7	0.5	594	15	US-10-127-901A-10 Sequence 10, Appl
686	7	0.5	594	15	US-10-128-693A-10 Sequence 10, Appl
687	7	0.5	594	15	US-10-131-813A-10 Sequence 10, Appl
688	7	0.5	594	15	US-10-131-818A-10 Sequence 10, Appl
689	7	0.5	594	15	US-10-131-823A-10 Sequence 10, Appl
690	7	0.5	594	15	US-10-131-824A-10 Sequence 10, Appl
691	7	0.5	594	15	US-10-131-830A-10 Sequence 10, Appl
692	7	0.5	594	15	US-10-131-837A-10 Sequence 10, Appl
693	7	0.5	594	15	US-10-137-872A-10 Sequence 10, Appl
694	7	0.5	594	15	US-10-147-500-10 Sequence 10, Appl
695	7	0.5	594	15	US-10-147-502-10 Sequence 10, Appl
696	7	0.5	594	15	US-10-147-515-10 Sequence 10, Appl
697	7	0.5	594	15	US-10-147-517-10 Sequence 10, Appl
698	7	0.5	594	15	US-10-147-526-10 Sequence 10, Appl
699	7	0.5	594	15	US-10-147-527-10 Sequence 10, Appl
700	7	0.5	594	15	US-10-121-041-10 Sequence 10, Appl
701	7	0.5	594	15	US-10-121-043-10 Sequence 10, Appl
702	7	0.5	594	15	US-10-121-047-10 Sequence 10, Appl
703	7	0.5	594	15	US-10-123-215-10 Sequence 10, Appl
704	7	0.5	594	15	US-10-123-902-10 Sequence 10, Appl
705	7	0.5	594	15	US-10-123-908-10 Sequence 10, Appl
706	7	0.5	594	15	US-10-123-909-10 Sequence 10, Appl
707	7	0.5	594	15	US-10-123-910-10 Sequence 10, Appl
708	7	0.5	594	15	US-10-124-813-10 Sequence 10, Appl
709	7	0.5	594	15	US-10-124-817-10 Sequence 10, Appl
710	7	0.5	594	15	US-10-125-922-10 Sequence 10, Appl
711	7	0.5	594	15	US-10-125-924-10 Sequence 10, Appl
712	7	0.5	594	15	US-10-140-860-10 Sequence 10, Appl
713	7	0.5	594	15	US-10-142-417-10 Sequence 10, Appl
714	7	0.5	594	15	US-10-147-519-10 Sequence 10, Appl
715	7	0.5	594	15	US-10-157-782-10 Sequence 10, Appl
716	7	0.5	594	15	US-10-152-395-10 Sequence 10, Appl
717	7	0.5	594	15	US-10-125-926A-10 Sequence 10, Appl
718	7	0.5	594	15	US-10-125-930A-10 Sequence 10, Appl
719	7	0.5	594	15	US-10-127-831A-10 Sequence 10, Appl
720	7	0.5	594	15	US-10-127-837A-10 Sequence 10, Appl
721	7	0.5	594	15	US-10-127-838B-10 Sequence 10, Appl
722	7	0.5	594	15	US-10-127-842A-10 Sequence 10, Appl
723	7	0.5	594	15	US-10-127-843A-10 Sequence 10, Appl
724	7	0.5	594	15	US-10-127-845A-10 Sequence 10, Appl
725	7	0.5	594	15	US-10-127-846A-10 Sequence 10, Appl
726	7	0.5	594	15	US-10-127-848A-10 Sequence 10, Appl
727	7	0.5	594	15	US-10-127-849A-10 Sequence 10, Appl
728	7	0.5	594	15	US-10-127-850A-10 Sequence 10, Appl
729	7	0.5	594	15	US-10-127-851A-10 Sequence 10, Appl
730	7	0.5	594	15	US-10-128-684A-10 Sequence 10, Appl
731	7	0.5	594	15	US-10-128-686A-10 Sequence 10, Appl
732	7	0.5	594	15	US-10-128-690A-10 Sequence 10, Appl
733	7	0.5	594	15	US-10-128-691A-10 Sequence 10, Appl
734	7	0.5	594	15	US-10-131-819A-10 Sequence 10, Appl
735	7	0.5	594	15	US-10-131-829A-10 Sequence 10, Appl
736	7	0.5	594	15	US-10-131-836A-10 Sequence 10, Appl
737	7	0.5	594	15	US-10-146-729-10 Sequence 10, Appl
738	7	0.5	594	15	US-10-146-791-10 Sequence 10, Appl
739	7	0.5	594	15	US-10-147-484-10 Sequence 10, Appl
740	7	0.5	594	15	US-10-147-508-10 Sequence 10, Appl
741	7	0.5	594	15	US-10-147-512-10 Sequence 10, Appl
742	7	0.5	594	15	US-10-175-735-10 Sequence 10, Appl
743	7	0.5	594	15	US-10-121-040-10 Sequence 10, Appl
744	7	0.5	594	15	US-10-121-056-10 Sequence 10, Appl
745	7	0.5	594	15	US-10-121-061-10 Sequence 10, Appl
746	7	0.5	594	15	US-10-123-235-10 Sequence 10, Appl
747	7	0.5	594	15	US-10-124-818-10 Sequence 10, Appl
748	7	0.5	594	15	US-10-137-868-10 Sequence 10, Appl
749	7	0.5	594	15	US-10-147-492-10 Sequence 10, Appl
750	7	0.5	594	15	US-10-158-782-10 Sequence 10, Appl
751	7	0.5	594	15	US-10-123-905-10 Sequence 10, Appl
752	7	0.5	594	15	US-10-123-907-10 Sequence 10, Appl
753	7	0.5	594	15	US-10-124-815-10 Sequence 10, Appl
754	7	0.5	594	15	US-10-125-921A-10 Sequence 10, Appl
755	7	0.5	594	15	US-10-125-928A-10 Sequence 10, Appl
756	7	0.5	594	15	US-10-127-821A-10 Sequence 10, Appl
757	7	0.5	594	15	US-10-127-822A-10 Sequence 10, Appl
758	7	0.5	594	15	US-10-127-824A-10 Sequence 10, Appl
759	7	0.5	594	15	US-10-127-826A-10 Sequence 10, Appl
760	7	0.5	594	15	US-10-127-827A-10 Sequence 10, Appl
761	7	0.5	594	15	US-10-127-828A-10 Sequence 10, Appl
762	7	0.5	594	15	US-10-127-830A-10 Sequence 10, Appl
763	7	0.5	594	15	US-10-127-832A-10 Sequence 10, Appl
764	7	0.5	594	15	US-10-127-833A-10 Sequence 10, Appl
765	7	0.5	594	15	US-10-127-834A-10 Sequence 10, Appl
766	7	0.5	594	15	US-10-127-836A-10 Sequence 10, Appl
767	7	0.5	594	15	US-10-127-841A-10 Sequence 10, Appl
768	7	0.5	594	15	US-10-127-844A-10 Sequence 10, Appl
769	7	0.5	594	15	US-10-128-687A-10 Sequence 10, Appl
770	7	0.5	594	15	US-10-128-688A-10 Sequence 10, Appl
771	7	0.5	594	15	US-10-128-689A-10 Sequence 10, Appl
772	7	0.5	594	15	US-10-128-694A-10 Sequence 10, Appl
773	7	0.5	594	15	US-10-131-825A-10 Sequence 10, Appl
774	7	0.5	594	15	US-10-230-417-10 Sequence 10, Appl
775	7	0.5	594	15	US-10-131-815A-10 Sequence 10, Appl
776	7	0.5	594	15	US-10-131-817A-10 Sequence 10, Appl
777	7	0.5	594	15	US-10-131-821A-10 Sequence 10, Appl
778	7	0.5	594	15	US-10-131-822A-10 Sequence 10, Appl
779	7	0.5	594	15	US-10-131-828A-10 Sequence 10, Appl
780	7	0.5	594	15	US-10-131-835A-10 Sequence 10, Appl
781	7	0.5	594	15	US-10-137-864A-10 Sequence 10, Appl
782	7	0.5	594	15	US-10-137-869A-10 Sequence 10, Appl
783	7	0.5	594	15	US-10-147-523-10 Sequence 10, Appl
784	7	0.5	594	15	US-10-158-785-10 Sequence 10, Appl
785	7	0.5	594	15	US-10-121-051-10 Sequence 10, Appl
786	7	0.5	594	15	US-10-121-042-10 Sequence 10, Appl
787	7	0.5	594	15	US-10-123-912-10 Sequence 10, Appl
788	7	0.5	594	15	US-10-192-007-10 Sequence 10, Appl
789	7	0.5	594	15	US-10-194-359-10 Sequence 10, Appl
790	7	0.5	594	15	US-10-127-847A-10 Sequence 10, Appl
791	7	0.5	594	15	US-10-175-590-10 Sequence 10, Appl
792	7	0.5	594	16	US-10-137-866-10 Sequence 10, Appl
793	7	0.5	594	16	US-10-146-726-10 Sequence 10, Appl
794	7	0.5	594	16	US-10-146-727-10 Sequence 10, Appl
795	7	0.5	594	16	US-10-146-788-10 Sequence 10, Appl
796	7	0.5	594	16	US-10-152-380-10 Sequence 10, Appl
797	7	0.5	594	16	US-10-153-934-10 Sequence 10, Appl
798	7	0.5	613	9	US-09-815-242-10990 Sequence 10990, A
799	7	0.5	635	12	US-09-848-841-16 Sequence 16, Appl
800	7	0.5	646	9	US-09-815-242-10311 Sequence 10311, A
801	7	0.5	646	9	US-09-815-242-13769 Sequence 13769, A
802	7	0.5	653	12	US-10-199-672-402 Sequence 402, App
803	7	0.5	653	12	US-10-187-749-402 Sequence 402, App
804	7	0.5	653	12	US-10-194-457-402 Sequence 402, App
805	7	0.5	653	12	US-10-184-642-402 Sequence 402, App
806	7	0.5	653	12	US-10-196-747-402 Sequence 402, App
807	7	0.5	653	12	US-10-173-689-402 Sequence 402, App
808	7	0.5	653	12	US-10-173-690-402 Sequence 402, App
809	7	0.5	653	12	US-10-173-691-402 Sequence 402, App
810	7	0.5	653	12	US-10-173-692-402 Sequence 402, App
811	7	0.5	653	12	US-10-173-694-402 Sequence 402, App
812	7	0.5	653	12	US-10-173-698-402 Sequence 402, App
813	7	0.5	653	12	US-10-173-699-402 Sequence 402, App
814	7	0.5	653	12	US-10-173-707-402 Sequence 402, App
815	7	0.5	653	12	US-10-174-569-402 Sequence 402, App
816	7	0.5	653	12	US-10-174-583-402 Sequence 402, App
817	7	0.5	653	12	US-10-174-587-402 Sequence 402, App
818	7	0.5	653	12	US-10-174-589-402 Sequence 402, App

819	7	0.5	653	12	US-10-174-591-402	Sequence 402, App
820	7	0.5	653	12	US-10-175-736-402	Sequence 402, App
821	7	0.5	653	12	US-10-175-742-402	Sequence 402, App
822	7	0.5	653	12	US-10-175-744-402	Sequence 402, App
823	7	0.5	653	12	US-10-175-745-402	Sequence 402, App
824	7	0.5	653	12	US-10-175-748-402	Sequence 402, App
825	7	0.5	653	12	US-10-175-751-402	Sequence 402, App
826	7	0.5	653	12	US-10-175-754-402	Sequence 402, App
827	7	0.5	653	12	US-10-176-480-402	Sequence 402, App
828	7	0.5	653	12	US-10-176-489-402	Sequence 402, App
829	7	0.5	653	12	US-10-176-754-402	Sequence 402, App
830	7	0.5	653	12	US-10-176-755-402	Sequence 402, App
831	7	0.5	653	12	US-10-176-759-402	Sequence 402, App
832	7	0.5	653	12	US-10-176-920-402	Sequence 402, App
833	7	0.5	653	12	US-10-176-922-402	Sequence 402, App
834	7	0.5	653	12	US-10-176-924-402	Sequence 402, App
835	7	0.5	653	12	US-10-176-984-402	Sequence 402, App
836	7	0.5	653	12	US-10-179-508-402	Sequence 402, App
837	7	0.5	653	12	US-10-179-512-402	Sequence 402, App
838	7	0.5	653	12	US-10-179-515-402	Sequence 402, App
839	7	0.5	653	12	US-10-173-702-402	Sequence 402, App
840	7	0.5	653	12	US-10-173-703-402	Sequence 402, App
841	7	0.5	653	12	US-10-173-704-402	Sequence 402, App
842	7	0.5	653	12	US-10-174-574-402	Sequence 402, App
843	7	0.5	653	12	US-10-176-486-402	Sequence 402, App
844	7	0.5	653	12	US-10-176-490-402	Sequence 402, App
845	7	0.5	653	12	US-10-176-752-402	Sequence 402, App
846	7	0.5	653	12	US-10-176-981-402	Sequence 402, App
847	7	0.5	653	12	US-10-176-983-402	Sequence 402, App
848	7	0.5	653	12	US-10-176-988-402	Sequence 402, App
849	7	0.5	653	12	US-10-179-517-402	Sequence 402, App
850	7	0.5	653	12	US-10-179-521-402	Sequence 402, App
851	7	0.5	653	12	US-10-202-475-402	Sequence 402, App
852	7	0.5	653	14	US-10-052-586-402	Sequence 402, App
853	7	0.5	653	15	US-10-174-590-402	Sequence 402, App
854	7	0.5	653	15	US-10-176-758-402	Sequence 402, App
855	7	0.5	653	15	US-10-175-737-402	Sequence 402, App
856	7	0.5	653	15	US-10-173-706-402	Sequence 402, App
857	7	0.5	653	15	US-10-175-738-402	Sequence 402, App
858	7	0.5	653	15	US-10-175-752-402	Sequence 402, App
859	7	0.5	653	15	US-10-176-482-402	Sequence 402, App
860	7	0.5	653	15	US-10-176-757-402	Sequence 402, App
861	7	0.5	653	15	US-10-176-913-402	Sequence 402, App
862	7	0.5	653	15	US-10-180-552-402	Sequence 402, App
863	7	0.5	653	15	US-10-180-557-402	Sequence 402, App
864	7	0.5	653	15	US-10-173-700-402	Sequence 402, App
865	7	0.5	653	15	US-10-174-572-402	Sequence 402, App
866	7	0.5	653	15	US-10-174-579-402	Sequence 402, App
867	7	0.5	653	15	US-10-174-582-402	Sequence 402, App
868	7	0.5	653	15	US-10-174-588-402	Sequence 402, App
869	7	0.5	653	15	US-10-175-739-402	Sequence 402, App
870	7	0.5	653	15	US-10-175-740-402	Sequence 402, App
871	7	0.5	653	15	US-10-175-743-402	Sequence 402, App
872	7	0.5	653	15	US-10-176-488-402	Sequence 402, App
873	7	0.5	653	15	US-10-176-492-402	Sequence 402, App
874	7	0.5	653	15	US-10-176-747-402	Sequence 402, App
875	7	0.5	653	15	US-10-176-750-402	Sequence 402, App
876	7	0.5	653	15	US-10-176-985-402	Sequence 402, App
877	7	0.5	653	15	US-10-176-987-402	Sequence 402, App
878	7	0.5	653	15	US-10-176-992-402	Sequence 402, App
879	7	0.5	653	15	US-10-176-993-402	Sequence 402, App
880	7	0.5	653	15	US-10-184-658-402	Sequence 402, App
881	7	0.5	653	15	US-10-176-991-402	Sequence 402, App
882	7	0.5	653	15	US-10-173-695-402	Sequence 402, App
883	7	0.5	653	15	US-10-173-697-402	Sequence 402, App
884	7	0.5	653	15	US-10-173-705-402	Sequence 402, App
885	7	0.5	653	15	US-10-174-576-402	Sequence 402, App
886	7	0.5	653	15	US-10-174-585-402	Sequence 402, App
887	7	0.5	653	15	US-10-174-586-402	Sequence 402, App
888	7	0.5	653	15	US-10-175-747-402	Sequence 402, App
889	7	0.5	653	15	US-10-176-481-402	Sequence 402, App
890	7	0.5	653	15	US-10-176-485-402	Sequence 402, App
891	7	0.5	653	15	US-10-176-487-402	Sequence 402, App

892	7	0.5	653	15	US-10-176-493-402	Sequence 402, App
893	7	0.5	653	15	US-10-176-756-402	Sequence 402, App
894	7	0.5	653	15	US-10-176-911-402	Sequence 402, App
895	7	0.5	653	15	US-10-176-919-402	Sequence 402, App
896	7	0.5	653	15	US-10-176-925-402	Sequence 402, App
897	7	0.5	653	15	US-10-176-978-402	Sequence 402, App
898	7	0.5	653	15	US-10-179-510-402	Sequence 402, App
899	7	0.5	653	15	US-10-180-543-402	Sequence 402, App
900	7	0.5	653	15	US-10-180-544-402	Sequence 402, App
901	7	0.5	653	15	US-10-180-546-402	Sequence 402, App
902	7	0.5	653	15	US-10-180-547-402	Sequence 402, App
903	7	0.5	653	15	US-10-180-549-402	Sequence 402, App
904	7	0.5	653	15	US-10-180-555-402	Sequence 402, App
905	7	0.5	653	15	US-10-180-559-402	Sequence 402, App
906	7	0.5	653	15	US-10-181-000-402	Sequence 402, App
907	7	0.5	653	15	US-10-183-010-402	Sequence 402, App
908	7	0.5	653	15	US-10-183-012-402	Sequence 402, App
909	7	0.5	653	15	US-10-184-614-402	Sequence 402, App
910	7	0.5	653	15	US-10-184-623-402	Sequence 402, App
911	7	0.5	653	15	US-10-184-635-402	Sequence 402, App
912	7	0.5	653	15	US-10-184-637-402	Sequence 402, App
913	7	0.5	653	15	US-10-184-646-402	Sequence 402, App
914	7	0.5	653	15	US-10-184-647-402	Sequence 402, App
915	7	0.5	653	15	US-10-184-652-402	Sequence 402, App
916	7	0.5	653	15	US-10-187-594-402	Sequence 402, App
917	7	0.5	653	15	US-10-187-596-402	Sequence 402, App
918	7	0.5	653	15	US-10-187-745-402	Sequence 402, App
919	7	0.5	653	15	US-10-187-885-402	Sequence 402, App
920	7	0.5	653	15	US-10-187-886-402	Sequence 402, App
921	7	0.5	653	15	US-10-199-464-402	Sequence 402, App
922	7	0.5	653	15	US-10-196-756-402	Sequence 402, App
923	7	0.5	653	15	US-10-176-751-402	Sequence 402, App
924	7	0.5	653	15	US-10-176-760-402	Sequence 402, App
925	7	0.5	653	15	US-10-176-990-402	Sequence 402, App
926	7	0.5	653	15	US-10-180-541-402	Sequence 402, App
927	7	0.5	653	15	US-10-180-542-402	Sequence 402, App
928	7	0.5	653	15	US-10-180-548-402	Sequence 402, App
929	7	0.5	653	15	US-10-180-551-402	Sequence 402, App
930	7	0.5	653	15	US-10-180-998-402	Sequence 402, App
931	7	0.5	653	15	US-10-180-999-402	Sequence 402, App
932	7	0.5	653	15	US-10-183-013-402	Sequence 402, App
933	7	0.5	653	15	US-10-184-612-402	Sequence 402, App
934	7	0.5	653	15	US-10-184-616-402	Sequence 402, App
935	7	0.5	653	15	US-10-184-617-402	Sequence 402, App
936	7	0.5	653	15	US-10-184-622-402	Sequence 402, App
937	7	0.5	653	15	US-10-184-628-402	Sequence 402, App
938	7	0.5	653	15	US-10-184-629-402	Sequence 402, App
939	7	0.5	653	15	US-10-184-630-402	Sequence 402, App
940	7	0.5	653	15	US-10-184-631-402	Sequence 402, App
941	7	0.5	653	15	US-10-184-632-402	Sequence 402, App
942	7	0.5	653	15	US-10-184-636-402	Sequence 402, App
943	7	0.5	653	15	US-10-184-640-402	Sequence 402, App
944	7	0.5	653	15	US-10-184-650-402	Sequence 402, App
945	7	0.5	653	15	US-10-184-651-402	Sequence 402, App
946	7	0.5	653	15	US-10-187-588-402	Sequence 402, App
947	7	0.5	653	15	US-10-187-597-402	Sequence 402, App
948	7	0.5	653	15	US-10-187-598-402	Sequence 402, App
949	7	0.5	653	15	US-10-187-600-402	Sequence 402, App
950	7	0.5	653	15	US-10-187-601-402	Sequence 402, App
951	7	0.5	653	15	US-10-187-602-402	Sequence 402, App
952	7	0.5	653	15	US-10-187-603-402	Sequence 402, App
953	7	0.5	653	15	US-10-187-741-402	Sequence 402, App
954	7	0.5	653	15	US-10-187-743-402	Sequence 402, App
955	7	0.5	653	15	US-10-187-746-402	Sequence 402, App
956	7	0.5	653	15	US-10-187-747-402	Sequence 402, App
957	7	0.5	653	15	US-10-187-751-402	Sequence 402, App
958	7	0.5	653	15	US-10-187-753-402	Sequence 402, App
959	7	0.5	653	15	US-10-187-754-402	Sequence 402, App
960	7	0.5	653	15	US-10-187-757-402	Sequence 402, App
961	7	0.5	653	15	US-10-187-884-402	Sequence 402, App
962	7	0.5	653	15	US-10-188-767-402	Sequence 402, App
963	7	0.5	653	15	US-10-188-769-402	Sequence 402, App
964	7	0.5	653	15	US-10-188-770-402	Sequence 402, App

```
 965      7    0.5    653  15  US-10-188-773-402         Sequence 402, App
 966      7    0.5    653  15  US-10-188-781-402         Sequence 402, App
 967      7    0.5    653  15  US-10-194-361-402         Sequence 402, App
 968      7    0.5    653  15  US-10-194-423-402         Sequence 402, App
 969      7    0.5    653  15  US-10-195-897-402         Sequence 402, App
 970      7    0.5    653  15  US-10-195-901-402         Sequence 402, App
 971      7    0.5    653  15  US-10-195-902-402         Sequence 402, App
 972      7    0.5    653  15  US-10-196-743-402         Sequence 402, App
 973      7    0.5    653  15  US-10-196-760-402         Sequence 402, App
 974      7    0.5    653  15  US-10-173-708-402         Sequence 402, App
 975      7    0.5    653  15  US-10-176-479-402         Sequence 402, App
 976      7    0.5    653  15  US-10-176-748-402         Sequence 402, App
 977      7    0.5    653  15  US-10-176-916-402         Sequence 402, App
 978      7    0.5    653  15  US-10-179-507-402         Sequence 402, App
 979      7    0.5    653  15  US-10-179-516-402         Sequence 402, App
 980      7    0.5    653  15  US-10-179-519-402         Sequence 402, App
 981      7    0.5    653  15  US-10-179-525-402         Sequence 402, App
 982      7    0.5    653  15  US-10-180-540-402         Sequence 402, App
 983      7    0.5    653  15  US-10-180-545-402         Sequence 402, App
 984      7    0.5    653  15  US-10-183-006-402         Sequence 402, App
 985      7    0.5    653  15  US-10-183-008-402         Sequence 402, App
 986      7    0.5    653  15  US-10-183-017-402         Sequence 402, App
 987      7    0.5    653  15  US-10-183-019-402         Sequence 402, App
 988      7    0.5    653  15  US-10-184-618-402         Sequence 402, App
 989      7    0.5    653  15  US-10-184-625-402         Sequence 402, App
 990      7    0.5    653  15  US-10-184-626-402         Sequence 402, App
 991      7    0.5    653  15  US-10-184-627-402         Sequence 402, App
 992      7    0.5    653  15  US-10-184-645-402         Sequence 402, App
 993      7    0.5    653  15  US-10-184-654-402         Sequence 402, App
 994      7    0.5    653  15  US-10-184-655-402         Sequence 402, App
 995      7    0.5    653  15  US-10-188-774-402         Sequence 402, App
 996      7    0.5    653  15  US-10-188-775-402         Sequence 402, App
 997      7    0.5    653  15  US-10-194-462-402         Sequence 402, App
 998      7    0.5    653  15  US-10-196-745-402         Sequence 402, App
 999      7    0.5    653  15  US-10-196-762-402         Sequence 402, App
1000      7    0.5    653  15  US-10-197-695-402         Sequence 402, App
```

ALIGNMENTS

```
RESULT 1
US-10-006-780-2
; Sequence 2, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
;  APPLICANT: Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
;  TITLE OF INVENTION:  METHODS FOR ITS USE
;  FILE REFERENCE: CYTOP083
;  CURRENT APPLICATION NUMBER: US/10/006,780
;  CURRENT FILING DATE:  2001-11-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1288
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-10-006-780-2

  Query Match             100.0%;  Score 1288;  DB 15;  Length 1288;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 60

Qy         61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120

Qy        121 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240

Qy        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300

Qy        301 GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGS 360

Qy        361 EMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKY 420

Qy        421 SSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNDNNDNNNDNNNNNNNDSS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 SSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNDNNDNNNDNNNNNNNDSS 480

Qy        481 SMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 540

Qy        541 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600

Qy        601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 660

Qy        661 NGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNK 720

Qy        721 NNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNS 780

Qy        781 HNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIH 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 HNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIH 840

Qy        841 GCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNE 900

Qy        901 MKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKN 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 MKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKN 960

Qy        961 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLD 1020

Qy       1021 DQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIK 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 DQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIK 1080

Qy       1081 NFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYIN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 NFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYIN 1140

Qy       1141 EDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 EDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN 1200
```

```
Qy       1201 NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSN 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSN 1260

Qy       1261 NGDIILLNKKLVQDNIKNSMDHNNIHKK 1288
              ||||||||||||||||||||||||||||
Db       1261 NGDIILLNKKLVQDNIKNSMDHNNIHKK 1288

RESULT 2
US-10-006-780-6
; Sequence 6, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
;  APPLICANT: Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Guo, Jun
;  APPLICANT:   Freedman, Richard
;  TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
;  TITLE OF INVENTION:  METHODS FOR ITS USE
;  FILE REFERENCE: CYTOP083
;  CURRENT APPLICATION NUMBER: US/10/006,780
;  CURRENT FILING DATE:  2001-11-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 355
;   TYPE: PRT
;   ORGANISM: P. Falciparum
US-10-006-780-6

  Query Match              27.5%;  Score 354;  DB 15;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  354;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 61

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 121

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 181

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 241

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 301

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDS 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDS 355

RESULT 3
US-10-006-780-4
; Sequence 4, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
;  APPLICANT: Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Guo, Jun
;  APPLICANT:   Freedman, Richard
;  TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
;  TITLE OF INVENTION:  METHODS FOR ITS USE
;  FILE REFERENCE: CYTOP083
;  CURRENT APPLICATION NUMBER: US/10/006,780
;  CURRENT FILING DATE:  2001-11-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 332
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-10-006-780-4

  Query Match              25.8%;  Score 332;  DB 15;  Length 332;
  Best Local Similarity   100.0%;  Pred. No. 2e-300;
  Matches  332;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 60

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 120

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 180

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 240

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 300

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFK 335
              ||||||||||||||||||||||||||||||||
Db        301 KSIMIANISPTISCCEQTLNTLRYSSRVKNFK 332

RESULT 4
US-10-006-780-10
; Sequence 10, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
;  APPLICANT: Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Guo, Jun
;  APPLICANT:   Freedman, Richard
;  TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
;  TITLE OF INVENTION:  METHODS FOR ITS USE
;  FILE REFERENCE: CYTOP083
;  CURRENT APPLICATION NUMBER: US/10/006,780
;  CURRENT FILING DATE:  2001-11-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 332
;   TYPE: PRT
;   ORGANISM: P. Falciparum
US-10-006-780-10

  Query Match              25.8%;  Score 332;  DB 15;  Length 332;
  Best Local Similarity   100.0%;  Pred. No. 2e-300;
  Matches  332;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 64
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 60

Qy         65 VDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAA 124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAA 120
```

```
Qy        125 GDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRV 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRV 180

Qy        185 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSE 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSE 240

Qy        245 RGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSK 304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSK 300

Qy        305 SIMIANISPTISCCEQTLNTLRYSSRVKNFKN 336
              ||||||||||||||||||||||||||||||||
Db        301 SIMIANISPTISCCEQTLNTLRYSSRVKNFKN 332
```

RESULT 5
US-10-006-780-8
; Sequence 8, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; APPLICANT: Guo, Jun
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
; TITLE OF INVENTION: METHODS FOR ITS USE
; FILE REFERENCE: CYTOP083
; CURRENT APPLICATION NUMBER: US/10/006,780
; CURRENT FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 361
; TYPE: PRT
; ORGANISM: P. Falciparum
US-10-006-780-8

```
  Query Match             25.6%;  Score 330;  DB 15;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-298;
  Matches  330;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 61

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 121

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 181

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 241

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 301

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKN 333
              ||||||||||||||||||||||||||||||
Db        302 KSIMIANISPTISCCEQTLNTLRYSSRVKN 331
```

RESULT 6
US-09-801-368-36
; Sequence 36, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-36

```
  Query Match             1.3%;  Score 17;  DB 10;  Length 666;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        351 NNNNNNNNNNNNNNNNI 367
```

RESULT 7
US-10-032-585-7714
; Sequence 7714, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7714
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7714

```
  Query Match             1.2%;  Score 16;  DB 12;  Length 213;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNN 756
              ||||||||||||||||
Db        172 TNNNNNNNNNNNNNNN 187
```

RESULT 8
US-09-994-485-8
; Sequence 8, Application US/09994485
; Patent No. US20020142429A1
; GENERAL INFORMATION:

```
;       APPLICANT: Ryazanov, Alexey G.
;                  Hait, William N.
;                  Pavur, Karen S.
;       TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)
;                           AND METHODS OF USE THEREFOR
;       NUMBER OF SEQUENCES: 25
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: David A. Jackson, Esq.
;            STREET: 411 Hackensack Ave, Continental Plaza, 4th
;                    Floor
;            CITY: Hackensack
;            STATE: New Jersey
;            COUNTRY: USA
;            ZIP: 07601
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/994,485
;            FILING DATE: 27-No. US20020142429A1-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/914,999
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Jackson Esq., David A.
;            REGISTRATION NUMBER: 26,742
;            REFERENCE/DOCKET NUMBER: 601-1-078
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 201-487-5800
;            TELEFAX: 201-343-1684
;  INFORMATION FOR SEQ ID NO: 8:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 732 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Dictyostelium discoideum
;       SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-994-485-8

  Query Match             1.2%;  Score 16;  DB 10;  Length 732;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        355 NNNNNNNNNNNNNNNN 370

RESULT 9
US-09-832-292-12
; Sequence 12, Application US/09832292
; Patent No. US20020177205A1
; GENERAL INFORMATION:
;  APPLICANT: Ryazanov, Alexey
;  TITLE OF INVENTION: MAMMALIAN ALPHA-KINASE PROTEINS, NUCLEIC ACIDS AND
;  TITLE OF INVENTION:  DIAGNOSTIC AND THERAPEUTIC USES THEREOF
;  FILE REFERENCE: 601-1-098CIP
;  CURRENT APPLICATION NUMBER: US/09/832,292
;  CURRENT FILING DATE:  2001-04-10
;  PRIOR APPLICATION NUMBER: 09/632,131
;  PRIOR FILING DATE: 2001-08-03
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 732
;   TYPE: PRT
;   ORGANISM: Dictyostelium discoideum
US-09-832-292-12

  Query Match             1.2%;  Score 16;  DB 10;  Length 732;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        355 NNNNNNNNNNNNNNNN 370

RESULT 10
US-09-801-368-224
; Sequence 224, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 224
;   LENGTH: 758
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-224

  Query Match             1.2%;  Score 16;  DB 10;  Length 758;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        605 NNNNNNNNNNNNNNNN 620

RESULT 11
US-10-032-585-7565
; Sequence 7565, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7565
;   LENGTH: 798
;   TYPE: PRT
```

```
;   ORGANISM: Candida albicans
US-10-032-585-7565

  Query Match             1.2%;  Score 16;  DB 12;  Length 798;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        390 NNNNNNNNNNNNNNNN 405


RESULT 12
US-09-801-368-132
; Sequence 132, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 132
;   LENGTH: 961
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-132

  Query Match             1.2%;  Score 16;  DB 10;  Length 961;
  Best Local Similarity   100.0%;  Pred. No. 6.7e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        515 NNNNNNNNNNNNNNNN 530


RESULT 13
US-10-087-464-50
; Sequence 50, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
;  APPLICANT: Chishti, Athar
;  APPLICANT:  Oh, Steven
;  APPLICANT:  Liu, David
;  APPLICANT:  Goel, Vikas
;  APPLICANT:  Li, Xuerong
;  TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
;  FILE REFERENCE: S1237/7019
;  CURRENT APPLICATION NUMBER: US/10/087,464
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 06/272,930
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 59
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 50
;   LENGTH: 1331
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-10-087-464-50

  Query Match             1.2%;  Score 16;  DB 15;  Length 1331;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db       1178 NNNNNNNNNNNNNNNN 1193


RESULT 14
US-10-135-322-17
; Sequence 17, Application US/10135322
; Publication No. US20020173017A1
; GENERAL INFORMATION:
;  APPLICANT: BENFEY, PN
;  APPLICANT:  HELARIUTTA, Y
;  APPLICANT:  MAHONEN, AP
;  APPLICANT:  BONKE, AWM
;  APPLICANT:  KAUPPINEN, L
;  APPLICANT:  RIIKONEN, M
;  TITLE OF INVENTION: WOODEN LEG GENE, PROMOTER AND USES THEREOF
;  FILE REFERENCE: 5914-086-999
;  CURRENT APPLICATION NUMBER: US/10/135,322
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: 60/253,739
;  PRIOR FILING DATE: 2000-11-29
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
;   LENGTH: 2150
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-135-322-17

  Query Match             1.2%;  Score 16;  DB 14;  Length 2150;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-05;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        105 NNNNNNNNNNNNNNNN 120


RESULT 15
US-10-304-095-6
; Sequence 6, Application US/10304095
; Publication No. US20030134275A1
; GENERAL INFORMATION:
;  APPLICANT: Long, David M.
;  APPLICANT:  Metz, Anneke M.
;  APPLICANT:  Love, Ruschelle A.
;  TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
;  FILE REFERENCE: 47714-5009-US
;  CURRENT APPLICATION NUMBER: US/10/304,095
;  CURRENT FILING DATE:  2002-11-26
;  PRIOR APPLICATION NUMBER: US/09/417,485
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 2184
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (330)..(335)
;   OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
```

```
;   OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
US-10-304-095-6

  Query Match              1.2%;  Score 16;  DB 12;  Length 2184;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-05;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        239 NNNNNNNNNNNNNNNN 254


RESULT 16
US-10-184-644-35
; Sequence 35, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C227
;  CURRENT APPLICATION NUMBER: US/10/184,644
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 35
;   LENGTH: 2690
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 2039-2065
;   OTHER INFORMATION: unknown base
US-10-184-644-35

  Query Match              1.2%;  Score 16;  DB 15;  Length 2690;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-05;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db       2039 NNNNNNNNNNNNNNNN 2054


RESULT 17
US-10-184-634-35
; Sequence 35, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C217
;  CURRENT APPLICATION NUMBER: US/10/184,634
;  CURRENT FILING DATE:  2002-06-28
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 35
;   LENGTH: 2690
;   TYPE: DNA
;   ORGANISM: Homo Sapien
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 2039-2065
;   OTHER INFORMATION: unknown base
US-10-184-634-35

  Query Match              1.2%;  Score 16;  DB 15;  Length 2690;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-05;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db       2039 NNNNNNNNNNNNNNNN 2054


RESULT 18
US-09-864-761-47535
; Sequence 47535, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
```

```
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47535
;   LENGTH: 73
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC010436.4
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  46
;   OTHER INFORMATION: EST_HUMAN HIT: BF327361.1, EVALUE 2.20e+00
US-09-864-761-47535

  Query Match             1.2%;  Score 15;  DB 9;  Length 73;
  Best Local Similarity  100.0%;  Pred. No. 5.8e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNN 756
              |||||||||||||||
Db         11 NNNNNNNNNNNNNNN 25


RESULT 19
US-10-422-866-4
; Sequence 4, Application US/10422866
; Publication No. US20030170857A1
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. US20030170857A1el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1054
;  CURRENT APPLICATION NUMBER: US/10/422,866
;  CURRENT FILING DATE:  2003-04-23
;  PRIOR APPLICATION NUMBER: US/09/722,129
;  PRIOR FILING DATE: 2000-11-24
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 303
;   TYPE: PRT
;   ORGANISM: Human
US-10-422-866-4

  Query Match             1.2%;  Score 15;  DB 12;  Length 303;
  Best Local Similarity  100.0%;  Pred. No. 2e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         91 CFAYGQTGSGKTYTM 105
              |||||||||||||||
Db         90 CFAYGQTGSGKTYTM 104


RESULT 20
US-10-422-866-2
; Sequence 2, Application US/10422866
; Publication No. US20030170857A1
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. US20030170857A1el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1054
;  CURRENT APPLICATION NUMBER: US/10/422,866
;  CURRENT FILING DATE:  2003-04-23
;  PRIOR APPLICATION NUMBER: US/09/722,129
;  PRIOR FILING DATE: 2000-11-24
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 492
;   TYPE: PRT
;   ORGANISM: Human
US-10-422-866-2

  Query Match             1.2%;  Score 15;  DB 12;  Length 492;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         91 CFAYGQTGSGKTYTM 105
              |||||||||||||||
Db        279 CFAYGQTGSGKTYTM 293


RESULT 21
US-10-304-095-8
; Sequence 8, Application US/10304095
; Publication No. US20030134275A1
; GENERAL INFORMATION:
;  APPLICANT: Long, David M.
;  APPLICANT:  Metz, Anneke M.
;  APPLICANT:  Love, Ruschelle A.
;  TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
;  FILE REFERENCE: 47714-5009-US
;  CURRENT APPLICATION NUMBER: US/10/304,095
;  CURRENT FILING DATE:  2002-11-26
;  PRIOR APPLICATION NUMBER: US/09/417,485
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 794
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-10-304-095-8

  Query Match             1.2%;  Score 15;  DB 12;  Length 794;
  Best Local Similarity  100.0%;  Pred. No. 4.8e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        744 NNNNNNNNNNNNNNI 758
              |||||||||||||||
Db        130 NNNNNNNNNNNNNNI 144


RESULT 22
US-09-826-752-6
; Sequence 6, Application US/09826752
; Patent No. US20010026930A1
; GENERAL INFORMATION:
;  APPLICANT: Guarente, Leonard P.
;  APPLICANT:  Austriaco Jr., Nicanor
;  APPLICANT:  Claus, James J.
;  APPLICANT:  Cole, Francesca
;  APPLICANT:  Kennedy, Brian
;  TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESCENCE IN
;  TITLE OF INVENTION:  YEAST
;  FILE REFERENCE: 0050.1491-005
;  CURRENT APPLICATION NUMBER: US/09/826,752
;  CURRENT FILING DATE:  2001-04-05
;  PRIOR APPLICATION NUMBER: US 08/396,001
;  PRIOR FILING DATE: 1995-02-28
;  PRIOR APPLICATION NUMBER: PCT/US94/09351
;  PRIOR FILING DATE: 1994-08-15
;  PRIOR APPLICATION NUMBER: US 08/107,408
;  PRIOR FILING DATE: 1993-08-16
;  PRIOR APPLICATION NUMBER: US 09/323,433
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 888
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-826-752-6

  Query Match             1.2%;  Score 15;  DB 9;  Length 888;
```

```
  Best Local Similarity   100.0%;  Pred. No. 5.3e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNN 756
              |||||||||||||||
Db        521 NNNNNNNNNNNNNNN 535


RESULT 23
US-09-839-185-2
; Sequence 2, Application US/09839185
; Patent No. US20020069433A1
;  GENERAL INFORMATION:
;       APPLICANT: Schmidt, Eduard Daniel Leendert
;                  De Vries, Sape Cornelis
;                  Hecht, Valerie France Gabrielle
;       TITLE OF INVENTION: Apomixis Conferred By Expression of SERK
;                           Interacting Proteins
;       NUMBER OF SEQUENCES: 18
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Syngenta Patent Dept.
;            STREET: 3054 Cornwallis Road
;            CITY: RTP
;            STATE: NC
;            COUNTRY: USA
;            ZIP: 27709
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/839,185
;            FILING DATE: 19-Apr-2001
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Meigs, J. Timothy
;            REGISTRATION NUMBER: 38,241
;            REFERENCE/DOCKET NUMBER: S-30683A
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 919-541-8587
;  INFORMATION FOR SEQ ID NO: 2:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 375 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Arabidopsis thaliana
;       IMMEDIATE SOURCE:
;            CLONE: 3A35
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-839-185-2

  Query Match             1.1%;  Score 14;  DB 9;  Length 375;
  Best Local Similarity   100.0%;  Pred. No. 0.00021;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNN 755
              ||||||||||||||
Db        262 NNNNNNNNNNNNNN 275


RESULT 24
US-10-032-585-7575
; Sequence 7575, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:. 2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7575
;   LENGTH: 458
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-10-032-585-7575

  Query Match             1.1%;  Score 14;  DB 12;  Length 458;
  Best Local Similarity   100.0%;  Pred. No. 0.00025;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNN 755
              ||||||||||||||
Db        187 NNNNNNNNNNNNNN 200


RESULT 25
US-10-195-144-37
; Sequence 37, Application US/10195144
; Publication No. US20030126646A1
; GENERAL INFORMATION:
;  APPLICANT: BROWN, GREGORY G.
;  APPLICANT:  FORMANOVA, NATASA
;  APPLICANT:  DENDY, CHARLES
;  APPLICANT:  LANDRY, BENOIT S.
;  APPLICANT:  CHEUNG, WING
;  APPLICANT:  JIN, HUA
;  TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN
;  TITLE OF INVENTION:  PLANTS
;  FILE REFERENCE: 16313-0136
;  CURRENT APPLICATION NUMBER: US/10/195,144
;  CURRENT FILING DATE:  2002-10-01
;  PRIOR APPLICATION NUMBER: 60/305,026
;  PRIOR FILING DATE: 2001-07-12
;  PRIOR APPLICATION NUMBER: 60/305,363
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: 60/308,736
;  PRIOR FILING DATE: 2001-07-30
;  NUMBER OF SEQ ID NOS: 128
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
;   LENGTH: 1062
;   TYPE: PRT
;   ORGANISM: Raphanus sativum
US-10-195-144-37

  Query Match             1.1%;  Score 14;  DB 15;  Length 1062;
  Best Local Similarity   100.0%;  Pred. No. 0.00053;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        523 FAYGQTGSGKTYTM 536


RESULT 26
US-09-801-368-298
; Sequence 298, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
```

```
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 298
;   LENGTH: 1770
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-298

  Query Match             1.1%;  Score 14;  DB 10;  Length 1770;
  Best Local Similarity   100.0%;  Pred. No. 0.00084;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNN 755
              ||||||||||||||
Db        345 NNNNNNNNNNNNNN 358


RESULT 27
US-10-032-585-7237
; Sequence 7237, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7237
;   LENGTH: 397
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-10-032-585-7237

  Query Match             1.0%;  Score 13;  DB 12;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 0.0019;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNN 754
              |||||||||||||
Db        177 NNNNNNNNNNNNN 189


RESULT 28
US-10-032-585-7930
; Sequence 7930, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7930
;   LENGTH: 941
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-10-032-585-7930

  Query Match             1.0%;  Score 13;  DB 12;  Length 941;
  Best Local Similarity   100.0%;  Pred. No. 0.0041;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNN 753
              |||||||||||||
Db        654 TNNNNNNNNNNNN 666


RESULT 29
US-09-801-368-134
; Sequence 134, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 134
;   LENGTH: 1151
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-134

  Query Match             1.0%;  Score 13;  DB 10;  Length 1151;
  Best Local Similarity   100.0%;  Pred. No. 0.0049;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 NNNNNNNNNNNNS 514
              |||||||||||||
Db         38 NNNNNNNNNNNNS 50


RESULT 30
US-10-060-019-31
; Sequence 31, Application US/10060019
; Publication No. US20030003564A1
; GENERAL INFORMATION:
;  APPLICANT: Tyers, Mike
;  APPLICANT:  Willems, Andrew
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
;  TITLE OF INVENTION:  DEPENDENT PROTEOLYSIS
;  FILE REFERENCE: 11757.10USU1
;  CURRENT APPLICATION NUMBER: US/10/060,019
```

```
;  CURRENT FILING DATE:  2002-01-29
;  PRIOR APPLICATION NUMBER: US/09/177,165
;  PRIOR FILING DATE: 1998-10-22
;  PRIOR APPLICATION NUMBER: 60/092,443
;  PRIOR FILING DATE: 1998-07-10
;  PRIOR APPLICATION NUMBER: 60/063,254
;  PRIOR FILING DATE: 1997-10-24
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
;   LENGTH: 1151
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-10-060-019-31

  Query Match              1.0%;  Score 13;  DB 15;  Length 1151;
  Best Local Similarity  100.0%;  Pred. No. 0.0049;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 NNNNNNNNNNNNS 514
              |||||||||||||
Db         38 NNNNNNNNNNNNS 50

RESULT 31
US-10-032-585-7034
; Sequence 7034, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7034
;   LENGTH: 387
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-10-032-585-7034

  Query Match              0.9%;  Score 12;  DB 12;  Length 387;
  Best Local Similarity  100.0%;  Pred. No. 0.016;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        499 INVNNNNNNNNN 510
              ||||||||||||
Db        238 INVNNNNNNNNN 249

RESULT 32
US-09-925-300-1074
; Sequence 1074, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1074
;   LENGTH: 410
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (14)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (21)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (177)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (248)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (300)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (372)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1074

  Query Match              0.9%;  Score 12;  DB 10;  Length 410;
  Best Local Similarity  100.0%;  Pred. No. 0.017;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         91 CFAYGQTGSGKT 102
              ||||||||||||
Db        143 CFAYGQTGSGKT 154

RESULT 33
US-10-032-585-7788
; Sequence 7788, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;  APPLICANT: Terry, Roemer D.
;  APPLICANT:  Bo, Jiang
;  APPLICANT:  Charles, Boone
;  APPLICANT:  Howard, Bussey
;  TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;  FILE REFERENCE: 10182-005-999
;  CURRENT APPLICATION NUMBER: US/10/032,585
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 8000
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7788
;   LENGTH: 511
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-10-032-585-7788

  Query Match              0.9%;  Score 12;  DB 12;  Length 511;
  Best Local Similarity  100.0%;  Pred. No. 0.02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        457 NNNNNNDNNDNN 468
              ||||||||||||
Db         50 NNNNNNDNNDNN 61

RESULT 34
US-09-849-602-20
; Sequence 20, Application US/09849602
; Publication No. US20030165834A1
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Matthew J.
;  APPLICANT:  Old, Lloyd J.
;  APPLICANT:  Stockert, Elisabeth
;  APPLICANT:  Chen, Yao-Tseng
;  TITLE OF INVENTION: Colon Cancer Antigen Panel
;  FILE REFERENCE: L0461/7105(JRV)
```

```
;  CURRENT APPLICATION NUMBER: US/09/849,602
;  CURRENT FILING DATE:  2001-05-04
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
;   LENGTH: 725
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-849-602-20

  Query Match             0.9%;  Score 12;  DB 12;  Length 725;
  Best Local Similarity   100.0%;  Pred. No. 0.028;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         91 CFAYGQTGSGKT 102
              ||||||||||||
Db        344 CFAYGQTGSGKT 355


RESULT 35
US-09-924-154-17
; Sequence 17, Application US/09924154
; Patent No. US20020127241A1
; GENERAL INFORMATION:
;  APPLICANT: Narum, David L.
;  APPLICANT:  Sim, Kim L.
;  TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
;  FILE REFERENCE: 05213-0465  43170-262105
;  CURRENT APPLICATION NUMBER: US/09/924,154
;  CURRENT FILING DATE:  2001-08-07
;  PRIOR APPLICATION NUMBER: US 60/223,525
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
;   LENGTH: 1501
;   TYPE: PRT
;   ORGANISM: Mammalian
US-09-924-154-17

  Query Match             0.9%;  Score 12;  DB 10;  Length 1501;
  Best Local Similarity   100.0%;  Pred. No. 0.052;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        503 NNNNNNNNNNNS 514
              ||||||||||||
Db         46 NNNNNNNNNNNS 57


RESULT 36
US-10-106-698-4938
; Sequence 4938, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
;  FILE REFERENCE: PA005P1
;  CURRENT APPLICATION NUMBER: US/10/106,698
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: PCT/US00/26524
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US 60/157,137
;  PRIOR FILING DATE: 1999-09-29
;  PRIOR APPLICATION NUMBER: US 60/163,280
;  PRIOR FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 8564
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4938
;   LENGTH: 160
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (3)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-4938

  Query Match             0.9%;  Score 11;  DB 15;  Length 160;
  Best Local Similarity   100.0%;  Pred. No. 0.061;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         93 AYGQTGSGKTY 103
              |||||||||||
Db         89 AYGQTGSGKTY 99


RESULT 37
US-09-934-455-462
; Sequence 462, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
;  APPLICANT: Adam, Luc
;  APPLICANT:  Creelman, Robert
;  APPLICANT:  Dubell, Arnold
;  APPLICANT:  Heard, Jacqueline
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Keddie, James
;  APPLICANT:  Pilgrim, Marsha
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Reuber, Lynne
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Pineda, Omaira
;  TITLE OF INVENTION: Genes for Modifying Plant Traits IV
;  FILE REFERENCE: MBI-0025
;  CURRENT APPLICATION NUMBER: US/09/934,455
;  CURRENT FILING DATE:  2001-08-22
;  PRIOR APPLICATION NUMBER: 60/227439
;  PRIOR FILING DATE: 2000-08-22
;  PRIOR APPLICATION NUMBER: MBI-0022
;  PRIOR FILING DATE: 2001-11-16
;  PRIOR APPLICATION NUMBER: MBI-0023
;  PRIOR FILING DATE: 2001-04-17
;  NUMBER OF SEQ ID NOS: 516
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 462
;   LENGTH: 250
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-934-455-462

  Query Match             0.9%;  Score 11;  DB 11;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 0.091;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 NNNNNNNNNNN 512
              |||||||||||
Db          4 NNNNNNNNNNN 14


RESULT 38
US-10-286-264-36
; Sequence 36, Application US/10286264
; Publication No. US20030093837A1
; GENERAL INFORMATION:
;  APPLICANT: Keddie, James
;  APPLICANT:  Riechmann, Jose-Luis
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Zhang, James
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Heard, Jacqueline
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Adam, Luc
;  APPLICANT:  Broun, Pierre
;  APPLICANT:  Reuber, Lynne
```

```
;  APPLICANT:  Pilgrim, Marsha
;  APPLICANT:  Samaha, Raymond
;  TITLE OF INVENTION: POLYNUCLEOTIDES FOR SEED TRAIT ALTERATION
;  FILE REFERENCE: MBI-008
;  CURRENT APPLICATION NUMBER: US/10/286,264
;  CURRENT FILING DATE:  2002-11-01
;  PRIOR APPLICATION NUMBER: 60/125,814
;  PRIOR FILING DATE: 1999-03-23
;  NUMBER OF SEQ ID NOS: 165
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
;   LENGTH: 298
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: G1362
US-10-286-264-36

  Query Match             0.9%;  Score 11;  DB 15;  Length 298;
  Best Local Similarity   100.0%;  Pred. No. 0.11;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        504 NNNNNNNNNNS 514
              |||||||||||
Db        204 NNNNNNNNNNS 214

RESULT 39
US-10-278-536-188
; Sequence 188, Application US/10278536
; Publication No. US20030131386A1
; GENERAL INFORMATION:
;  APPLICANT: Samaha, Raymond
;  APPLICANT:  Heard, Jacqueline
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Reuber, Lynne
;  APPLICANT:  Riechmann, Jose-Luis
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Keddie, James
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Pilgrim, Marsha
;  APPLICANT:  Adam, Luc
;  APPLICANT:  Broun, Pierre
;  TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
;  FILE REFERENCE: MBI-011
;  CURRENT APPLICATION NUMBER: US/10/278,536
;  CURRENT FILING DATE:  2002-10-22
;  PRIOR APPLICATION NUMBER: 60/125,814
;  PRIOR FILING DATE: 1999-03-23
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 188
;   LENGTH: 298
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   OTHER INFORMATION: G1362
US-10-278-536-188

  Query Match             0.9%;  Score 11;  DB 16;  Length 298;
  Best Local Similarity   100.0%;  Pred. No. 0.11;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        504 NNNNNNNNNNS 514
              |||||||||||
Db        204 NNNNNNNNNNS 214

RESULT 40
US-09-991-496-134
; Sequence 134, Application US/09991496
; Patent No. US20020169285A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Webb, John R.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Bhatia, Ajay
;  APPLICANT:  Coler, Rhea
;  APPLICANT:  Probst, Peter
;  APPLICANT:  Brannon, Mark
;  TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
;  TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;  FILE REFERENCE: 210121.420C9
;  CURRENT APPLICATION NUMBER: US/09/991,496
;  CURRENT FILING DATE:  2001-11-20
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 134
;   LENGTH: 309
;   TYPE: PRT
;   ORGANISM: Leishmania major and chagasi
US-09-991-496-134

  Query Match             0.9%;  Score 11;  DB 10;  Length 309;
  Best Local Similarity   100.0%;  Pred. No. 0.11;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKT 102
              |||||||||||
Db        256 FAYGQTGSGKT 266

Search completed: October  2, 2003, 16:45:20
Job time : 102 secs
```

```
OM protein - protein search, using sw model

Run on:         October  2, 2003, 16:07:05 ; Search time 29 Seconds
                                             (without alignments)
                                             2088.634 Million cell updates/sec

Title:          US-10-006-780-2
Perfect score:  6842
Sequence:       1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	761.5	11.1	730	1	KCM1_XENLA	Q91636 xenopus lae
2	752.5	11.0	725	1	KF2C_HUMAN	Q99661 homo sapien
3	747.5	10.9	671	1	KF2C_MACFA	Q951p1 macaca fasc
4	735.5	10.7	679	1	KIF2_HUMAN	O00139 homo sapien
5	735	10.7	718	1	KF2C_CRIGR	P70096 cricetulus
6	732	10.7	721	1	KF2C_MOUSE	Q922s8 mus musculu
7	728	10.6	716	1	KIF2_MOUSE	P28740 mus musculu
8	722	10.6	671	1	KF2C_RAT	Q62909 rattus norv
9	719.5	10.5	682	1	KIF2_XENLA	Q91637 xenopus lae
10	597	8.7	537	1	ARP_PLAFA	P04931 plasmodium
11	576	8.4	624	1	DSK1_CYLFU	Q39493 cylindrothe
12	561.5	8.2	1111	1	KIP1_YEAST	P28742 saccharomyc
13	560.5	8.2	2339	1	RPC1_PLAFA	P27625 plasmodium
14	543	7.9	1956	1	ATX1_PLAFA	Q04956 plasmodium
15	538.5	7.9	1085	1	CUT7_SCHPO	P24339 schizosacch
16	520.5	7.6	1056	1	K125_ARATH	P82266 arabidopsis
17	513	7.5	1006	1	K125_TOBAC	O23826 nicotiana t
18	500.5	7.3	967	1	KINH_LOLPE	P21613 loligo peal
19	500	7.3	1184	1	BIMC_EMENI	P17120 emericella
20	482.5	7.1	2663	1	CENE_HUMAN	Q02224 homo sapien
21	477	7.0	784	1	KL68_DROME	P46867 drosophila
22	472.5	6.9	747	1	KF3B_HUMAN	O15066 homo sapien
23	472.5	6.9	747	1	KF3B_MOUSE	Q61771 mus musculu
24	465.5	6.8	554	1	KLP3_SCHPO	Q9us60 schizosacch
25	465.5	6.8	1585	1	P3K3_DICDI	P54675 dictyosteli
26	463	6.8	1038	1	CIN8_YEAST	P27895 saccharomyc
27	460	6.7	702	1	KF3A_HUMAN	Q9y496 homo sapien
28	459	6.7	1066	1	KL61_DROME	P46863 drosophila
29	453	6.6	963	1	KINH_HUMAN	P33176 homo sapien
30	452.5	6.6	805	1	KIP3_YEAST	P53086 saccharomyc
31	449	6.6	883	1	KLP5_SCHPO	O14343 schizosacch
32	447.5	6.5	1057	1	KF11_HUMAN	P52732 homo sapien
33	447	6.5	963	1	KINH_MOUSE	Q61768 mus musculu
34	446	6.5	1648	1	KF14_HUMAN	Q15058 homo sapien
35	445.5	6.5	701	1	KF3A_MOUSE	P28741 mus musculu
36	443.5	6.5	956	1	KF5C_MOUSE	P28738 mus musculu
37	442.5	6.5	1226	1	KF4A_XENLA	Q91784 xenopus lae
38	439.5	6.4	1067	1	EG52_XENLA	Q91783 xenopus lae
39	439	6.4	957	1	KF5C_HUMAN	O60282 homo sapien
40	439	6.4	975	1	KINH_DROME	P17210 drosophila
41	437.5	6.4	1231	1	KF4A_MOUSE	P33174 mus musculu
42	435	6.4	786	1	FL10_CHLRE	P46869 chlamydomon
43	435	6.4	989	1	PTP3_DICDI	P54637 dictyosteli
44	434	6.3	2452	1	RPB1_PLAFD	P14248 plasmodium
45	433.5	6.3	742	1	KI21_STRPU	P46871 strongyloce

ALIGNMENTS

```
RESULT 1
KCM1_XENLA
ID   KCM1_XENLA     STANDARD;      PRT;   730 AA.
AC   Q91636;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin central motor 1 (XKCM1).
GN   KCM1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=96140638; PubMed=8548824;
RA   Walczak C.E., Mitchison T.J., Desai A.;
RT   "XKCM1: a Xenopus kinesin-related protein that regulates microtubule
RT   dynamics during mitotic spindle assembly.";
RL   Cell 84:37-47(1996).
RN   [2]
RP   REVISION TO 516.
RC   TISSUE=Ovary;
RA   Walczak C.E.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REGULATES MICROTUBULE DYNAMICS DURING MITOTIC SPINDLE
CC       ASSEMBLY. REQUIRED FOR BOTH ESTABLISHMENT AND MAINTENANCE OF
CC       MITOTIC SPINDLES. MAY ACT DIRECTLY ON MICROTUBULES, CAUSING A
CC       DESTABILIZATION AND EVENTUAL DEPOLYMERIZATION OF THE MICROTUBULE.
CC   -!- SUBUNIT: Homodimer (Potential).
CC   -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U36485; AAC59743.2; -.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    256       GLOBULAR (POTENTIAL).
FT   DOMAIN      257    598       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      599    730       COILED COIL (POTENTIAL).
```

```
FT   NP_BIND     352    359       ATP (POTENTIAL).
SQ   SEQUENCE   730 AA;  82585 MW;  25C31C187E491523 CRC64;

  Query Match             11.1%;  Score 761.5;  DB 1;  Length 730;
  Best Local Similarity   42.1%;  Pred. No. 1.3e-26;
  Matches  175;  Conservative   85;  Mismatches  129;  Indels   27;  Gaps   12;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |  ||: |||:| :   : : ||: |||:|||:|   |  |  ||:
Db        262 RICVCVRKRPLNKQELSKKEIDIISVPSKNIVLVHEPKLKVDLTKYLENQAFRFDFSFDE 321

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |  |  ||  | :||:  ::| |   :||||||||||||:|| |      |: : |:: :
Db        322 TATNEVVYRFTARPLVQSIFEGG-KATCFAYGQTGSGKTHTMGGDFSGKSQNVSKGVYAF 380

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
              |: |:|  |   |:   |     :|::|:||| ||::||| |:  :  ||: |:|| |
Db        381 ASRDVFLLL---DQPRYKHLDLDVFVTFFEIYNGKVFDLLNKKTKLRVLEDAKQEVQVVG 437

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSL-GKIA 236
              |   :|:: :: : |||: |   |  |    |  |||||| | | |:   : : | || :
Db        438 LLEKQVISADD-VFKMIEIGSACRTSGQTFANTSSSRSHACLQIILR---RGSKLHGKFS 493

Qy        237 FIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLR 296
               :||||:||| || | :: |: :|| ||||||||||||||:  :|:| |||:|:||::||
Db        494 LVDLAGNERGVDTASADRITRMEGAEINRSLLALKECIRALGQNKSHTPFRESKLTQILR 553

Qy        297 DIFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISIL 355
              | |:|: |:: ||| :||  : || |||||||: |||    ::       :|| : : :
Db        554 DSFIGENSRTCMIAMLSPGFNSCEYTLNTLRYADRVKELSPQNA------ETNDDNLQME 607

Qy        356 DSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
              || ||    :||| : :: : || :   :  |     | : |   |::|: |: ||
Db        608 DSGGSH---ASIEGLQLQDDFLLKDEELSTHN--SFQDALNRVGELEDKAVDELRE 658


RESULT 2
KF2C_HUMAN
ID   KF2C_HUMAN     STANDARD;      PRT;   725 AA.
AC   Q99661; Q96C18; Q96HB8; Q9BWV8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin-like protein KIF2C (Mitotic centromere-associated kinesin)
DE   (MCAK) (Kinesin-like protein 6).
GN   KIF2C OR KNSL6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=T-cell;
RX   MEDLINE=98094213; PubMed=9434124;
RA   Kim I.-G., Jun D.Y., Sohn U., Kim Y.H.;
RT   "Cloning and expression of human mitotic centromere-associated kinesin
RT   gene.";
RL   Biochim. Biophys. Acta 1359:181-186(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Testis;
RA   Cheng L.J.;
RT   "Cloning of human testis specific mitotic centromere-associated
RT   kinesin.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Kidney, and Uterus;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Present throughout the cell cycle, associates with
CC       centromeres at early prophase, and remains associated with the
CC       centromere until after telophase (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q99661-1; Sequence=Displayed;
CC       Name=2; Synonyms=TS-MCAK;
CC         IsoId=Q99661-2; Sequence=VSP_002866;
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THYMUS AND TESTIS,
CC       AT LOW LEVELS IN SMALL INTESTINE, THE MUCOSAL LINING OF COLON, AND
CC       PLACENTA, AND AT VERY LOW LEVELS IN SPLEEN AND OVARY; EXPRESSION
CC       IS NOT DETECTED IN PROSTATE, PERIPHERAL BLOOD LEUKOCYTES, HEART,
CC       BRAIN, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, OR PANCREAS.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U63743; AAC27660.1; -.
DR   EMBL; AY026505; AAK20168.1; -.
DR   EMBL; BC014924; AAH14924.1; -.
DR   EMBL; BC008764; AAH08764.1; -.
DR   HSSP; P17119; 3KAR.
DR   Genew; HGNC:6393; KIF2C.
DR   MIM; 604538; -.
DR   GO; GO:0005871; C:kinesin complex; TAS.
DR   GO; GO:0003777; F:microtubule motor activity; TAS.
DR   GO; GO:0008283; P:cell proliferation; TAS.
DR   GO; GO:0007002; P:centromere binding; TAS.
DR   GO; GO:0007067; P:mitosis; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil;
KW   Nuclear protein; Alternative splicing.
FT   DOMAIN        1    254       GLOBULAR (POTENTIAL).
FT   DOMAIN      255    518       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      618    658       COILED COIL (POTENTIAL).
FT   NP_BIND     348    355       ATP (POTENTIAL).
FT   DOMAIN      415    418       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   VARSPLIC      1     55       MAMDSSLQARLFPGLAIKIQRSNGLIHSANVRTVNLEKSCV
FT                                SVEWAEGGATKGKE -> M (in isoform 2).
FT                                /FTId=VSP_002866.
FT   CONFLICT    449    449       I -> L (IN REF. 1 AND 2).
FT   CONFLICT    698    698       R -> P (IN REF. 3; AAH08764).
SQ   SEQUENCE   725 AA;  81312 MW;  5BDECC133AB4B55C CRC64;
```

```
  Query Match             11.0%;  Score 752.5;  DB 1;  Length 725;
  Best Local Similarity   42.1%;  Pred. No. 3.1e-26;
  Matches  174;  Conservative   75;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |   ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db        258 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 317

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
               |  |  ||  | :||:  ::| |   :|||||||||||:|| |        |: : ||:
Db        318 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 376

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |    | |    : ::::|:|||  |||:|||  |:   :   ||:||::|  |   |:
Db        377 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 435

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                | : :: ::||||  |    |    |   | |||||    ||  |:     |      || : :|
Db        436 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 492

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||  | ::||: :||  ||:|||||||||||:   :|  |  |||:|:||:|||| |
Db        493 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 552

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || ||||||: |||          |        |:              ::  :
Db        553 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 603

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                   || |  |    |      :|                    |: : :     |: |: :: :|
Db        604 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 654
```

```
RESULT 3
KF2C_MACFA
ID   KF2C_MACFA     STANDARD;      PRT;   671 AA.
AC   Q95LP1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein KIF2C (Mitotic centromere-associated kinesin)
DE   (MCAK).
GN   KIF2C.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA   Terao K., Sugano S.;
RT   "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT   libraries.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Present throughout the cell cycle, associates with
CC       centromeres at early prophase, and remains associated with the
CC       centromere until after telophase (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB072747; BAB69716.1; -.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil;
KW   Nuclear protein.
FT   DOMAIN        1    200       GLOBULAR (POTENTIAL).
FT   DOMAIN      201    464       KINESIN-MOTOR.
FT   DOMAIN      564    604       COILED COIL (POTENTIAL).
FT   NP_BIND     294    301       ATP (POTENTIAL).
SQ   SEQUENCE   671 AA;  75639 MW;  4EAB399E8B6BBA46 CRC64;
```

```
  Query Match             10.9%;  Score 747.5;  DB 1;  Length 671;
  Best Local Similarity   41.9%;  Pred. No. 4.7e-26;
  Matches  173;  Conservative   76;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |   ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db        204 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 263

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
               |  |  ||  | :||:  ::| |   :|||||||||||:|| |        |: : ||:
Db        264 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 322

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |    | |    : ::::|:|||  |||:|||  |:   :   ||:||::|  |   |:
Db        323 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 381

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                | : :: ::||||  |    |    |   | |||||    ||  |:     |      || : :|
Db        382 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 438

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||  | ::||: :||  ||:|||||||||||:   :|  |  |||:|:||:|||| |
Db        439 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 498

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || ||||||: |||          |        |:              ::  :
Db        499 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 549

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                   || |  |    |      :|                    |: : :     |: :: :: :|
Db        550 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEERAVEELKE 600
```

```
RESULT 4
KIF2_HUMAN
ID   KIF2_HUMAN     STANDARD;      PRT;   679 AA.
AC   O00139;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Kinesin-like protein KIF2 (Kinesin-2) (HK2).
GN   KIF2 OR KNS2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97321046; PubMed=9177777;
RA   Debernardi S., Fontanella E., de Gregorio L., Pierotti M.A.,
RA   Delia D.;
RT   "Identification of a novel human kinesin-related gene (HK2) by the
RT   cDNA differential display technique.";
RL   Genomics 42:67-73(1997).
CC   -!- FUNCTION: THE KINESIN FAMILY MAY RECOGNIZE, TRANSPORT AND POSITION
CC       THEIR SPECIFIC CARGOES IN A SINGLE TYPE OF NEURONAL CELL.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y08319; CAA69621.1; -.
DR   HSSP; P17119; 3KAR.
DR   Genew; HGNC:6318; KIF2.
DR   MIM; 602591; -.
DR   GO; GO:0004002; F:adenosinetriphosphatase activity; TAS.
DR   GO; GO:0003774; F:motor activity; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Neurone.
FT   DOMAIN        1    190       GLOBULAR (POTENTIAL).
FT   DOMAIN      191    525       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      525    679       COILED COIL (POTENTIAL).
FT   NP_BIND     286    293       ATP (BY SIMILARITY).
SQ   SEQUENCE   679 AA;  76927 MW;  F96A88BA871BECED CRC64;

  Query Match             10.7%;  Score 735.5;  DB 1;  Length 679;
  Best Local Similarity   45.8%;  Pred. No. 1.6e-25;
  Matches  167;  Conservative   62;  Mismatches  111;  Indels   25;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | |||||:: | :  || |:||: :    : : ||: |||:|:|:|   |  |  |||
Db        196 RICVCVRKRPLNKKETQMKDLDVITIPSKDVVMVHEPKQKVDLTRYLENQTFRFDYAFDD 255

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
               :  |  ||  | |||:  ::| | : :|||||||||||:|| |        |  : ||:
Db        256 SAPNEMVYRFTAKPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 314

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
              || |:|  |    | | |       :: :|:|||  ||::|||  ::  :   ||:||::| |
Db        315 AARDVFLMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 371

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237
              |:   |   |: ::||:|| |   |   |  ||     ||||||:  | |:    |      ||:
Db        372 LQEREVKCVED-VLKLIDIGNSCRTSGQTSANAHSSRSHAVFQIILR--RKGKLHGKFSL 428

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297
              |||||:||||||  | ::||: :|| ||:|||||||||||:   :| |  |||  |:||:||||
Db        429 IDLAGNERGADTSSADRQTRLEGAEINKSLLALKECIRALGRNKPHTPFRASKLTQVLRD 488

Qy        298 IFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCI-----------NEED 345
                |:|: |:: ||| ||| :: || |||||||::|||        |                 |: |
Db        489 SFIGENSRTCMIATISPGMASCENTLNTLRYANRVKELTVDPTAAGDVRPIMHHPPNQID 548

Qy        346 DTNTE 350
              |  |:
Db        549 DLETQ 553

RESULT 5
KF2C_CRIGR
ID   KF2C_CRIGR     STANDARD;      PRT;   718 AA.
AC   P70096;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein KIF2C (Mitotic centromere-associated kinesin)
DE   (MCAK) (Kinesin-like protein 6).
GN   KIF2C OR KNSL6.
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10029;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95122643; PubMed=7822426;
RA   Wordeman L., Mitchison T.J.;
RT   "Identification and partial characterization of mitotic centromere-
RT   associated kinesin, a kinesin-related protein that associates with
RT   centromeres during mitosis.";
RL   J. Cell Biol. 128:95-105(1995).
RN   [2]
RP   REVISIONS.
RA   Wordeman L.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Present throughout the cell cycle, associates with
CC       centromeres at early prophase, and remains associated with the
CC       centromere until after telophase.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U11790; AAB17358.2; -.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil;
KW   Nuclear protein.
FT   DOMAIN        1    248       GLOBULAR (POTENTIAL).
FT   DOMAIN      249    512       KINESIN-MOTOR.
FT   DOMAIN      613    651       COILED COIL (POTENTIAL).
FT   DOMAIN      689    716       COILED COIL (POTENTIAL).
FT   NP_BIND     342    349       ATP (POTENTIAL).
FT   DOMAIN      409    412       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ   SEQUENCE   718 AA;  80918 MW;  16ABD8BC66AD11B2 CRC64;

  Query Match             10.7%;  Score 735;  DB 1;  Length 718;
  Best Local Similarity   40.9%;  Pred. No. 1.8e-25;
  Matches  172;  Conservative   75;  Mismatches  136;  Indels   38;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | |||||:: |  ||: |:|:| : | |:: ||: |||:|||:|   |  |  ||:
Db        252 RICVCVRKRPLNKQELAKKEIDVISVPSKCLLFVHEPKLKVDLTKYLENQAFCFDFAFDE 311

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
               |  |  ||  | :||:  ::| |   :|||||||||||:|| |        |: : ||:
Db        312 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKSQNTSKGIYAM 370

Qy        123 AAGDIFTF--------LNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEV 174
              |: |:|         ||:          ::::|:||| ||::||| |:  :  ||: |::|
Db        371 ASRDVFLLKSQPRYRNLNL-------EVYVTFFEIYNGKVFDLLNKKAKLRVLEDSKQQV 423

Qy        175 VVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGK 234
               | |:  |   :::|  :  |   |   |   |   ||||||  | |:   |    ||
Db        424 QVVGLQEYLVNCADDVIKMLNMGSACRTSGQTFANSNSSRSHACFQILLR--AKGRLHGK 481

Qy        235 IAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKV 294
               : :||||:|||||| | ::||: :|| ||:|||||||||||:  :| | |||:|:||:|
Db        482 FSLVDLAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQV 541

Qy        295 LRDIFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERIS 353
              ||| |:|: |:: ||| ||| || || |||||||: |||         |      |:
```

```
Db        542 LRDSFIGENSRTCMIAMISPGISSCEYTLNTLRYADRVKELSPHSGLSGEQP-------- 593

Qy        354 ILDSKGSEMNASSIEN---VVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPR 410
               :  :  || |||       |  |    ||:  ::       |: : :    |: :: :: |
Db        594 -IQMETEEMEASSNGTSLAVNFKEEEELSSQMSS------FNEAMSQIRELEERAMEELR 646

Qy        411 E 411
              |
Db        647 E 647


RESULT 6
KF2C_MOUSE
ID   KF2C_MOUSE     STANDARD;      PRT;   721 AA.
AC   Q922S8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin-like protein KIF2C (Mitotic centromere-associated kinesin)
DE   (MCAK).
GN   KIF2C.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Present throughout the cell cycle, associates with
CC       centromeres at early prophase, and remains associated with the
CC       centromere until after telophase (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; BC006841; AAH06841.1; -.
DR   MGD; MGI:1921054; Kif2c.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil;
KW   Nuclear protein.
FT   DOMAIN        1    250       GLOBULAR (POTENTIAL).
FT   DOMAIN      251    514       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      614    652       COILED COIL (POTENTIAL).
FT   NP_BIND     348    355       ATP (POTENTIAL).
SQ   SEQUENCE   721 AA;  81085 MW;  92FFBFFDA05B7E35 CRC64;

  Query Match             10.7%;  Score 732;  DB 1;  Length 721;
  Best Local Similarity   41.2%;  Pred. No. 2.4e-25;
  Matches  170;  Conservative   78;  Mismatches  141;  Indels   24;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :|  | ||||||::  |    ||:  |:|:|  :  |  |   :  ||:  |||:|||:|    |    |  ||:
Db        254 RICVCVRKRPLNKQELAKKEIDVISVPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 313

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |   |   ||   |  :||:   ::| |    :||||||||||||:|| |      |:  :  ||:
Db        314 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKSQNASKGIYAM 372

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |:  |:|     |      |     |  :  ::::|:|||  ||::|||  |:   :    ||:  :::|  |   |:
Db        373 ASRDVFLLKNQPRYRNLNLE-VYVTFFEIYNGKVFDLLNKKAKLRVLEDSRQQVQVVGLQ 431

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                 |   :: ::|||: |     |    |      |   |||||    |  |:    |     || : :|
Db        432 EYLVTCADD-VIKMINMGSACRTSGQTFANSNSSRSHACFQILLR--TKGRLHGKFSLVD 488

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||  |  ::||:  :||  ||:|||||||||||:   :|  |  |||:|:||:||||  |
Db        489 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 548

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|:  |::  |||  |||  ||  ||  |||||||:  |||          |       |:           :     |
Db        549 IGENSRTCMIAMISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQPVQMETEVMEASSN 608

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
              |:  :   |  :    |:  :  |                |: :  :    |: :: :: ||
Db        609 GTSLTGNEEEEL---SSQMSS-----------FNEAMTQIRELEERALEELRE 647


RESULT 7
KIF2_MOUSE
ID   KIF2_MOUSE     STANDARD;      PRT;   716 AA.
AC   P28740;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin-like protein KIF2.
GN   KIF2 OR KIF2A OR KNS2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93077686; PubMed=1447303;
RA   Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
RA   Hirokawa N.;
RT   "Kinesin family in murine central nervous system.";
RL   J. Cell Biol. 119:1287-1296(1992).
CC   -!- FUNCTION: THE KINESIN FAMILY MAY RECOGNIZE, TRANSPORT AND POSITION
CC       THEIR SPECIFIC CARGOES IN A SINGLE TYPE OF NEURONAL CELL.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AMONG VARIOUS TISSUES UBIQUITOUSLY;
CC       ITS EXPRESSION LEVEL IN BRAIN DECREASES WITH DEVELOPMENT.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D12644; BAA02165.1; -.
DR   PIR; A44259; A44259.
DR   HSSP; P17119; 3KAR.
DR   MGD; MGI:108390; Kif2a.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Neurone.
FT   DOMAIN        1    189       GLOBULAR (POTENTIAL).
FT   DOMAIN      190    539       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      540    716       COILED COIL (POTENTIAL).
FT   NP_BIND     285    292       ATP (BY SIMILARITY).
SQ   SEQUENCE   716 AA;  80946 MW;  6653D01F00138A1B CRC64;

  Query Match             10.6%;  Score 728;  DB 1;  Length 716;
  Best Local Similarity   47.6%;  Pred. No. 3.6e-25;
  Matches  161;  Conservative   61;  Mismatches  102;  Indels   14;  Gaps    8;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63

Db        195 RICVCVRKRPLNKKETQMKDLDVITIPSKDVVMVHEPKQKVDLTRYLENQTFRFDYAFDD 254

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122

Db        255 SAPNEMVYRFTARPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 313

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178

Db        314 AARDVFLMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 370

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237

Db        371 LQEREVKCVED-VLKLIDIGNSCRTSGQTSANAHSSRSHAVFQIILR--RKGKLHGKFSL 427

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297

Db        428 IDLAGNERGADTSSADRQTRLEGAEINKSLLRLKECIRALGRNKPHTPFRASKLTQVLRD 487

Qy        298 IFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNF 334

Db        488 SFIGENSRTCMIATISPGMASCENTLNTLRYANRVKEF 525

RESULT 8
KF2C_RAT
ID   KF2C_RAT       STANDARD;      PRT;   671 AA.
AC   Q62909;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Kinesin-like protein KIF2C (Mitotic centromere-associated kinesin)
DE   (MCAK) (Kinesin-related protein 2).
GN   KIF2C OR KRP2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Testis;
RX   MEDLINE=96228687; PubMed=8688559;
RA   Sperry A.O., Zhao L.-P.;
RT   "Kinesin-related proteins in the mammalian testes: candidate motors
RT   for meiosis and morphogenesis.";
RL   Mol. Biol. Cell 7:289-305(1996).
RN   [2]
RP   REVISIONS.
RA   Sperry A.O.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Present throughout the cell cycle, associates with
CC       centromeres at early prophase, and remains associated with the
CC       centromere until after telophase (By similarity). Active in
CC       meiosis.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: TESTIS. LOCALIZED TO THE MEIOTICALLY ACTIVE
CC       CELLS OF THE SEMINIFEROUS EPITHELIA IN THE TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U44979; AAC53528.1; -.
DR   PIR; T10755; T10755.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil;
KW   Nuclear protein; Meiosis.
FT   DOMAIN        1    200       GLOBULAR (POTENTIAL).
FT   DOMAIN      201    464       KINESIN-MOTOR.
FT   DOMAIN      564    669       COILED COIL (POTENTIAL).
FT   NP_BIND     294    301       ATP (POTENTIAL).
SQ   SEQUENCE   671 AA;  75661 MW;  F2B54598C78DE8DE CRC64;

  Query Match             10.6%;  Score 722;  DB 1;  Length 671;
  Best Local Similarity   41.2%;  Pred. No. 6.1e-25;
  Matches  171;  Conservative   78;  Mismatches  138;  Indels   28;  Gaps   11;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63

Db        204 RICVCVRKRPLNKQELAKKEIDVISVPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 263

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122

Db        264 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKSQNASKGIYAM 322

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180

Db        323 ASRDVFLLKNQPRYRSLNLE-VYVTFFEIYNGKVFELLNKKAKLRVLEDSKQQVQVVGLQ 381

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239

Db        382 EYLVTCADD-VIKMINMGSACRTSGQTFANSNSSRSHACFQILLR--AKGRLHGKFSLVD 438

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299

Db        439 LAGNERGADTSSADRQTRMEGAEINKSLLALKESIRALGQNKAHTPFRESKLTQVLRDSF 498

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358

Db        499 IGENSRTCMIAMISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------AVQME 549

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKIN--RGKINDKIERNNILKNKSFDKPRE 411

Db        550 TEEMDASS-------HGASLTGNEEEELSSQMSSFNEAMTQIRELEERAMEELRE 597

RESULT 9
KIF2_XENLA
ID   KIF2_XENLA     STANDARD;      PRT;   682 AA.
AC   Q91637;
```

```
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin-like protein KIF2 (Kinesin-related protein XKIF2).
GN   KIF2.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A., AND REVISIONS.
RC   TISSUE=Ovary;
RA   Walczak C.E.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 124-682 FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=96140638; PubMed=8548824;
RA   Walczak C.E., Mitchison T.J., Desai A.;
RT   "XKCM1: a Xenopus kinesin-related protein that regulates microtubule
RT   dynamics during mitotic spindle assembly.";
RL   Cell 84:37-47(1996).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U36486; AAC59744.2; ALT_INIT.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    192       GLOBULAR (POTENTIAL).
FT   DOMAIN      193    542       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      543    682       COILED COIL (POTENTIAL).
FT   NP_BIND     288    295       ATP (BY SIMILARITY).
SQ   SEQUENCE   682 AA;  77414 MW;  3DA295BB319063F8 CRC64;

  Query Match             10.5%;  Score 719.5;  DB 1;  Length 682;
  Best Local Similarity   47.1%;  Pred. No. 7.9e-25;
  Matches  162;  Conservative   62;  Mismatches  105;  Indels   15;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
               :|  | ||||||::  |     || |:||:       : : ||: |||:|:::|    |   |   ||:
Db        198 RICVCVRKRPLNKKETTIKDLDVITIPIKDVVMVHEPKQKVDLTRFLENQTFRFDYAFDE 257

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
               |   |  | |||   |  :||:  ::|  |  : :||||||||||||:||  |        |   :  ||:
Db        258 TAPNETVYRFTARPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 316

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
               || |:|    |        |  |  |        :: :|:|||  ||::|||  ::     :   ||:||::| |
Db        317 AARDVFQMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 373

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237
               |:    |     |:  :||:|:  |      |    |    |   |  ||||||:    |  |:        |    || :
Db        374 LQEREVKCVED-VLKLIEIGNSCRTSGQTSANAHSSRSHAVFQIILR--KKGKMHGKFSL 430

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297
               |||||:||||||  |  ::|:|:  :||  ||:||||||||||||:    :|   |   |||  |:||:||||
Db        431 IDLAGNERGADTSSADRQTRLEGAEINKSLLALKECIRALGRNKPHTPFRASKLTQVLRD 490

Qy        298 IFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTC 340
               |:|:  |::  |||  |||  ::  ||  |||||||::||||       :  |   |
Db        491 SFIGENSRTCMIATISPGMASCENTLNTLRYANRVKEL-DPSRC 533

RESULT 10
ARP_PLAFA
ID   ARP_PLAFA      STANDARD;      PRT;   537 AA.
AC   P04931;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Asparagine-rich protein (AG319) (ARP) (Fragment).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86176787; PubMed=2421257;
RA   Stahl H.-D., Bianco A.E., Crewther P.E., Burkot T., Coppel R.L.,
RA   Brown G.V., Anders R.F., Kemp D.J.;
RT   "An asparagine-rich protein from blood stages of Plasmodium
RT   falciparum shares determinants with sporozoites.";
RL   Nucleic Acids Res. 14:3089-3102(1986).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M24328; AAA29491.1; -.
DR   PIR; A23770; A23770.
KW   Malaria.
FT   NON_TER       1      1
FT   NON_TER     537    537
SQ   SEQUENCE   537 AA;  63000 MW;  A6D76B15318CC239 CRC64;

  Query Match              8.7%;  Score 597;  DB 1;  Length 537;
  Best Local Similarity   28.5%;  Pred. No. 1.3e-19;
  Matches  212;  Conservative   92;  Mismatches  193;  Indels  246;  Gaps   39;

Qy        424 NDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMV 483
               |: :|     |       |  |||::|   |:      |||    |   |     |     ||:      |  |  ||:|     |
Db          2 NNNNKNNNNNDDGNINYQNT--NEFKDNKK--NMNFKNQYNNNYKFDENMNNSNTMHSRN 57

Qy        484 NNMINHMINNNIN-NNINVNN------------NNNNNNNNNSHNNHLPQPNYAFTDTS 530
               :|:   |:  ||:|:  || ||:||                       |  |  |  |  |  |
Db         58 SNVEEHLRNNSIDMNNSNINNYTNQQTRFSSFMENENENENKNYHTG------------- 104

Qy        531 DFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKL 590
                          :|||          :|  ||    ||       |  ||    |   |
Db        105 ----------GMNNN-----IHFKNKYDN----NNSSMKNTDNNK--------------- 130

Qy        591 TMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNV-----LLGLNKNTHH 645
               | ||:                |    ||     ||     | |        |::   |:              :|||
Db        131 TDTSYN---------MKGTIN--NDNNNMDYLRNINNINEYKGSAKNKFYTNYMNKNNLK 179

Qy        646 DISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMN-SNSI-YKSNYNSNQS 703
                        ::|  |  |:  ||          ||:|  |:  ||      ::     |:|||  :|||     |  |  |:|  :
Db        180 FTQNNNDNMNINEDNN------NNNNNNNNNNGVFSNYQNNNMNRNNSINIKRNLNNNNN 233

Qy        704 ISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENY 763
               |::         :|:|  : :||  ::                 :||  |   |         |     |:  |||  |||
Db        234 INN----NMNKMGSQDKNQNS--------NNNFYMNYNYQNRKNSMNNNMNNN------- 274

Qy        764 NNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDK 823
                          ||:|                  :||:        :||  |  :   |:|  |   ||       |:|::||              |:
Db        275 -----MNNNM------NHNM---NNNMNHNMNNNMNHNMNN---NMNHNMN------NNM 311

Qy        824 NYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMK 883
```

```
                 |                 ||:  |:::  | :   |    | ||||       :|  | |  :|
Db        312 N-------------NINSLDSDMSPNYHAHVKMSMMN-YNNN-----ESNTANPN--QMN 350

Qy        884 YGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTF 943
              :      |: ||||     :|| | :      || :  |||  |:    |      ||
Db        351 F----EQTNNDNMKR---ENNNMNNYGYDDNTVHVNNNTPSTDFFSRAVGYN-------- 395

Qy        944 QNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLST 1003
               |:| :|:|            |:|| :   |                           |
Db        396 -NNYLNNNN----------NMNSAVNNNS---------------------------SN 415

Qy       1004 NEDIYNKEMEGKHIRLDDQDKYDDND--NNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNN 1061
               :: |:    | |::           |:||  ||| :||  |:: :::|||  ||  ||: ||
Db        416 GNNMKNENSENKNVA-------DNNDSLNNNKNNNNNINMNESINNNNTLNN--NNEYNN 466

Qy       1062 VDNNNVDNDDDD---------VDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSL-DISSM 1111
                |||  |  ||||   .        :|  ::|             :|||  |:|     | |:|
Db        467 -QNNNEDEDDDDWGELGEDKYIDINSIMK---------KKNV--ILNQLEADLNDLSKK 513

Qy       1112 YDDTKEILNNILLSKYKAEKDNV 1134
               :| |         | | :||::
Db        514 GNDGKN------KKKNKMKKDDL 530
```

RESULT 11
DSK1_CYLFU

```
ID   DSK1_CYLFU     STANDARD;      PRT;   624 AA.
AC   Q39493;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Diatom spindle kinesin 1.
GN   DSK1.
OS   Cylindrotheca fusiformis (Marine diatom).
OC   Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae;
OC   Bacillariophycidae; Bacillariales; Bacillariaceae; Cylindrotheca.
OX   NCBI_TaxID=2853;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96222363; PubMed=8636234;
RA   Wein H., Foss M., Brady B., Cande W.Z.;
RT   "DSK1, a novel kinesin-related protein from the diatom Cylindrotheca
RT   fusiformis that is involved in anaphase spindle elongation.";
RL   J. Cell Biol. 133:595-604(1996).
CC   -!- FUNCTION: INVOLVED IN ANAPHASE SPINDLE ELONGATION.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U51680; AAB05681.1; -.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1     85       GLOBULAR (POTENTIAL).
FT   DOMAIN       86    425       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      426    624       COILED COIL (POTENTIAL).
FT   NP_BIND     186    193       ATP (POTENTIAL).
SQ   SEQUENCE   624 AA;  70308 MW;  DC996BFEC22BE155 CRC64;

  Query Match             8.4%;  Score 576;  DB 1;  Length 624;
  Best Local Similarity   37.5%;  Pred. No. 1.3e-18;
  Matches  138;  Conservative   60;  Mismatches  138;  Indels   32;  Gaps    6;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVD-MTKYIERHEFIVDKV 60
              || | : |||||:|: |::| | | ::    |  ::|    : ||| :|||:  : | :|
Db         93 NSNICIAVRKRPISDKERQKLDHDSVSCFQN-KVWIHSAKLKVDGITKYLTHNSFQLDHT 151

Qy         61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120
              | :       :|  |   ||:  :         : |  ||||||||||| | |
Db        152 FGEDSTTEQIYLATTLPLVDHVVSTQGRATVFCYGQTGSGKTYTMNGIQ----------- 200

Qy        121 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
               |  | |::  |   :   |      | ::|:|:|:| | : |||   :        ||:|  || :  |:
Db        201 QILAYDLYGQLAEHTDD--LEITVAFFELYSGNVLDLLHGCQRCKLLEDGNGEVNITGLR 258

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
                :  | |  :  : :|  ||           || ||||||  : |:|   |     ||:  :||
Db        259 EVPAPTPEAFLQVIEEGHSLRTTQKTEANDASSRSHAICQVFLRDYGGNLR-GKLGLVDL 317

Qy        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
              ||||||:||    | |  :|: |:|| ||||||||||:      |:|:| |:|| :|:| |
Db        318 AGSERGSDTKQHNSQRRTESADINTSLLALKECIRALGQKSAHVPYRGSKLTLILKDCFS 377

Qy        301 GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGS 360
                 ||: |:| :||  |   : :||||||: |:|                  :|:|     :|
Db        378 PDSKTTMVATVSPGASAADHSLNTLRYADRIKE----------------QRVSSNGQRGK 421

Qy        361 EMNASSIE 368
                 ||: |
Db        422 AAKASNRE 429
```

RESULT 12
KIP1_YEAST

```
ID   KIP1_YEAST     STANDARD;      PRT;  1111 AA.
AC   P28742;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Kinesin-like protein KIP1.
GN   KIP1 OR CIN9 OR YBL063W OR YBL0504 OR YBL0521.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92317166; PubMed=1618910;
RA   Roof D.M., Meluh P.B., Rose M.D.;
RT   "Kinesin-related proteins required for assembly of the mitotic
RT   spindle.";
RL   J. Cell Biol. 118:95-108(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=94205266; PubMed=8154187;
RA   Scherens B., el Bakkoury M., Vierendeels F., Dubois E., Messenguy F.;
RT   "Sequencing and functional analysis of a 32,560 bp segment on the
RT   left arm of yeast chromosome II. Identification of 26 open reading
RT   frames, including the KIP1 and SEC17 genes.";
RL   Yeast 9:1355-1371(1993).
RN   [3]
RP   CHARACTERIZATION.
RC   STRAIN=S288c;
RX   MEDLINE=92354062; PubMed=1643659;
RA   Saunders W.S., Hoyt M.A.;
RT   "Kinesin-related proteins required for structural integrity of the
RT   mitotic spindle.";
RL   Cell 70:451-458(1992).
CC   -!- FUNCTION: REQUIRED FOR ASSEMBLY OF THE MITOTIC SPINDLE. INTERACT
CC       WITH SPINDLE MICROTUBULES TO PRODUCE AN OUTWARDLY DIRECTED
CC       FORCE ACTING UPON THE POLES. FOLLOWING SPINDLE ASSEMBLY, CIN8 AND
CC       KIP1 APPARENTLY ACT TO OPPOSE A FORCE THAT DRAWS SEPARATED POLES
```

```
CC        BACK TOGETHER. THIS FORCE SEEMS TO BE MEDIATE BY KAR3.
CC     -!- SUBUNIT: MIGHT BE DIMERIC.
CC     -!- SUBCELLULAR LOCATION: SPINDLE MICROTUBULES THAT LIE BETWEEN THE
CC        POLES.
CC     -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC        SUBFAMILY.
CC     --------------------------------------------------------------------------
CC     This SWISS-PROT entry is copyright. It is produced through a collaboration
CC     between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC     the European Bioinformatics Institute.  There are no  restrictions on  its
CC     use  by  non-profit  institutions as long  as its content  is  in  no  way
CC     modified and this statement is not removed.  Usage  by  and for commercial
CC     entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC     or send an email to license@isb-sib.ch).
CC     --------------------------------------------------------------------------
DR     EMBL; Z11962; CAA78019.1; -.
DR     EMBL; Z23261; CAA80785.1; -.
DR     EMBL; Z35824; CAA84883.1; -.
DR     PIR; A42640; A42640.
DR     HSSP; P17119; 3KAR.
DR     SGD; S0000159; KIP1.
DR     GO; GO:0005816; C:spindle pole body; IDA.
DR     GO; GO:0005200; F:structural constituent of cytoskeleton; IPI.
DR     GO; GO:0007020; P:microtubule nucleation; IPI.
DR     GO; GO:0000092; P:mitotic anaphase B; IMP.
DR     InterPro; IPR001752; kinesin_motor.
DR     Pfam; PF00225; kinesin; 1.
DR     PRINTS; PR00380; KINESINHEAVY.
DR     SMART; SM00129; KISc; 1.
DR     PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR     PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW     Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW     Mitosis; Cell cycle.
FT     DOMAIN       50    423       KINESIN-MOTOR (BY SIMILARITY).
FT     DOMAIN      424    510       COILED COIL (POTENTIAL).
FT     DOMAIN      648    670       COILED COIL (POTENTIAL).
FT     DOMAIN      710    780       COILED COIL (POTENTIAL).
FT     DOMAIN      808    828       COILED COIL (POTENTIAL).
FT     NP_BIND     141    148       ATP (BY SIMILARITY).
SQ     SEQUENCE   1111 AA;  125794 MW;  212F8279766137FC CRC64;

  Query Match             8.2%;  Score 561.5;  DB 1;  Length 1111;
  Best Local Similarity   22.9%;  Pred. No. 1e-17;
  Matches  255;  Conservative  204;  Mismatches  393;  Indels  263;  Gaps   48;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITV---KNNCTLYIDEPRYKVDMTKYIERHEFIVD 58
               :| | | || |  :: | ::| | |:      :       :  :    :|  : ::|  |
Db         50 DSNIHVYVRCRSRNKREIEEKSSVVISTLGPQGKEIILSNGSHQSYSSSK--KTYQF--D 105

Qy         59 KVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLG---------S 108
              :||    |  ||:  | |   | ::  :|  |: ||||||:||||||  |
Db        106 QVFGAESDQETVFNATAKNYIKEML-HGYNCTIFAYGQTGTGKTYTMSGDINILGDVQST 164

Qy        109 QPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQ--------- 159
                   :  ||     |:|  |:  :|: :  : ||| |:|   | |||
Db        165 DNLLLGEHAGIIPRVLVDLFKELSSLNKEYS--VKISFLELYNENLKDLLSDSEDDDPAV 222

Qy        160 ---KRKM-VAALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRS 215
                 ||:: :    |     ::|| :: : : :   | :  :: | | ||:     || ||||
Db        223 NDPKRQIRIFDNNNNSSIMVKGMQEIFINSAHEGLNLLMQGSLKRKVAATKCNDLSSRS 282

Qy        216 HAILNI-------DLKD--INKN-TSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINR 265
              | :  |      | ||   |||    :||:  :|||||     : ::||:  |   |  ||:
Db        283 HTVFTITTNIVEQDSKDHGQNKNFVKIGKLNLVDLAGSENINRSGAENKRAQEAGL-INK 341

Qy        266 SLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTL 325
              ||| |      | |:      ||||:|:|:|:|::|:|     | :|: :|| |||      |:| :||
Db        342 SLLTLGRVINALVDHSNHIPYRESKLTRLLQDSLGGMTKTCIIATISPAKISMEETASTL 401

Qy        326 RYSSRVKNFKN---------KSTCINE-----------EDDTNTERISILDSKGSEMNA 364
               |::| |: ||          | ||: :               ::  | : | |     :::
Db        402 EYATRAKSIKNTPQVNQSLSKDTCLKDYIQEIEKLRNDLKNSRNKQGIFITQ---DQLDL 458

Qy        365 SSIENVVIKSNHLLSNNNNNKINRGKIN--DKIERNNILKN---------KSFDKPREGF 413
                 :::|    :|  :|      :|  : | |   :::: ||:|::              ::|:      |
Db        459 YESNSILIDEQNLKIHNLREQIKKFKENYLNQLDINNLLQSEKEKLIAIIQNFNVDFSNF 518

Qy        414 TSTFGK--YSSLNDIDKIKKNKKKGLINYKSTLYNDN-------------TINKKHNNNN 458
                 |    |  :::|   :::: : :         | |  :   || |          |:|      : |
Db        519 YSEIQKIHHTNLELMNEVIQQRDFSLENSQKQ-YNTNQNMQLKISQQVLQTLNTLQGSLN 577

Qy        459 NNNNDNND---------NNNDNNNNNNNDSS-SMVNNMINHMINNNINNNI-NVNNNNNN 507
               | |:   ::              | | :     :||:     : |:  ::: | :|    : ::: :
Db        578 NYNSKCSEVIKGVTEELTRNVNTHKAKHDSTLKSLLNITTNLLMNQMNELVRSISTSLEI 637

Qy        508 NNNNNNSHN----NHLPQPNYAF-------------TDTSDFSSLDDM--NCHLNNNDK 547
                 ::: ||      | : |  :   |                 :|   :|::::   ||   |   |
Db        638 FQSDSTSHYRKDLNEIYQSHQQFLKNLQNDIKSCLDSIGSSILTSINEISQNCTTNLNSM 697

Qy        548 SIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNL----------------------HLAR- 583
               :: :  :     :  :| :    |   :  | ||:|                        |::|
Db        698 NVLIENQQSGSSKLIKEQ---DLEIKKLKNDLINERRISNQFNQQLAEMKRYFQDHVSRT 754

Qy        584 ----HSVGSKLTMFSYDPQKNKDNTFFKSN----------INKMEDNTPKDILYESRN-- 627
                   |  :|       |  |   |   |   :       :||:  ::  :    :  :   |
Db        755 RSEFHDELNKCIDNLKDKQSKLDQDIWQKTASIFNETDIVVNKIHSDSIASLAHNAENTL 814

Qy        628 --VSNMN---GNVLLGLNKNTHHDISTK----DENHNDNKINNGVINIINNSNVNSINNS 678
                 ||    |      | |: |::  : |||:|       |     |||:   :      :  |   : :|
Db        815 KTVSQNNESFTNDLISLSRGMNMDISSKLRSLPINEFLNKISQTICETCGDDNTIA-SNP 873

Qy        679 NMNSINN-SNMNSNSIYKSNYNSNQSISDVQIRYVNEMDT-SNKNNDNIFFDAISCDNNM 736
                : ||     |:  :  |  :|          |  ::|     ::::|  ||:|| |:
Db        874 VLTSIKKFQNIICSDIALTNEKIMSLIDEIQ----SQIETISNENNINL----------- 918

Qy        737 YPNITNNNNNNNNNN----NNNNNNIDVENYNNRDGTNNSMKLYA--------YNSHNLF 784
                | | | |:  |     :  :  |  : :      :    :  : ||  :         : :|::
Db        919 ---IAINENFNSLCNFILTDYDENIMQISKTQDEVLSEHCEKLQSLKILGMDIFTAHSIE 975

Qy        785 QP---DNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKD-KNIH 840
                :|       |  :         ::|      |  |          | ||  :  |   |
Db        976 KPLHEHTRPEASVIKALPL--------LDYPKQFQIYR--------DAENKSKDDTSNSR 1019

Qy        841 GCDNNIIQNRN----DFEKK----------------KKTNFYNNNNIVIVNNNMG----- 875
                 |  |:  ||    |   |                   |   |   :|    :  |  |
Db       1020 TCIPNLSTNENFPLSQFSPKTPVPVPDQPLPKVLIPKSINSAKSNRSKTLPNTEGTGRES 1079

Qy        876 NNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNE 910
               ||  |            |:   :|   |:||::  |:
Db       1080 QNNLKR-------RFTTEPILKGEETENNDILQNK 1107


RESULT 13
RPC1_PLAFA
ID   RPC1_PLAFA     STANDARD;      PRT;  2339 AA.
AC   P27625;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   DNA-directed RNA polymerase III largest subunit (EC 2.7.7.6).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92018020; PubMed=1656254;
RA   Li W.B., Bzik D.J., Tanaka M., Gu H., Fox B.A., Inselburg J.;
RT   "Characterization of the gene encoding the largest subunit of
RT   Plasmodium falciparum RNA polymerase III.";
RL   Mol. Biochem. Parasitol. 46:229-240(1991).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT
```

```
CC       SUBUNITS. THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE
CC       III.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC       FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC       PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC       III FOR 5S AND TRNA GENES.
CC   -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M73770; AAA29729.1; -.
DR   InterPro; IPR000722; RNA_pol_A.
DR   InterPro; IPR007080; RNA_pol_Rpb1_1.
DR   InterPro; IPR007066; RNA_pol_Rpb1_3.
DR   InterPro; IPR007083; RNA_pol_Rpb1_4.
DR   InterPro; IPR007081; RNA_pol_Rpb1_5.
DR   InterPro; IPR006592; RNA_polA_N.
DR   Pfam; PF04997; RNA_pol_Rpb1_1; 1.
DR   Pfam; PF00623; RNA_pol_Rpb1_2; 1.
DR   Pfam; PF04983; RNA_pol_Rpb1_3; 1.
DR   Pfam; PF05000; RNA_pol_Rpb1_4; 1.
DR   Pfam; PF04998; RNA_pol_Rpb1_5; 1.
DR   SMART; SM00663; RPOLA_N; 1.
KW   Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
KW   Zinc-finger; Nuclear protein.
FT   ZN_FING      88    101       C3H-TYPE (POTENTIAL).
SQ   SEQUENCE   2339 AA;  272829 MW;  EDD899363086FD48 CRC64;

  Query Match              8.2%;  Score 560.5;  DB 1;  Length 2339;
  Best Local Similarity   20.3%;  Pred. No. 2.5e-17;
  Matches  356;  Conservative  234;  Mismatches  520;  Indels  645;  Gaps   73;

Qy         31 NNCTLYIDEPRYKVDMTKYIERHEFIVDKVF-----DDTVDNFTVYE-------NTIKPL 78
Db        684 SDASLYIELPIPAILKPKELWTGKQLISVLIKPNKKENTIINFEIQEREYSNKFGDLKHL 743

Qy         79 IID-----LYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY-----AAGDIF 128
Db        744 CLNDSYVCFYKSELIC-----GSLGK----KVLGSSKY------GLFYYLIHHNSSHIAL 788

Qy        129 TFLNIYDK-----DNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVV---KDLK 180
Db        789 KIMNRLSKLTSRYFSNKGMTIGIDDVRPSQTLTEKKKDLLLKGYEKVNNEIILYNEKKMQ 848

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
Db        849 IQPGCTLEETL-----EIKVKSILDDLRNDAGKTCNQYLHYLNKPLIMFNSGAKGALINI 903

Qy        241 A------GSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSE---- 290
Db        904 AQMIACVGQQNVAGQRIQN-------GFINRTL------------PHFHFHCKDSESRGF 944

Qy        291 ----------------------------------------LTKVLRDIFVGKSKSIMIA 309
Db        945 VQNSFYTGLSPTEFFFHTMSGREGLVDTAVKTAETGYMQRRLMKALEDLSIHYDYSVR-- 1002

Qy        310 NISPTISCCEQTLNTLRYSSRVKN---FKNKSTCINEEDDTNTERISILDS--------- 357
Db       1003 ------SCDKQIVQFI-YGDDALNPSYIDNNNTYLDQFDKVFDHIVSISSSHLLLSYKNK 1055

Qy        358 -----------KGSEMNASSIENVVIKSNHLLSNNNNNKINRGKI-NDKIERNNILKNKS 405
Db       1056 IPYLPHVQHQNKTSNMNN-------IYNN--MNNINNNDSNRSIIYNNDSNMNNINNNDS 1106

Qy        406 FDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLY-NDNTINKKHNNNNNNNNDN 464
Db       1107 ------NMNSIHNNNSNMNNIHNNDSN--------RSIIHNNDSNMNSIHNNDSNMNSIH 1152

Qy        465 NDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHN-------- 516
Db       1153 NNNSNMNNIHNNDSNRSIIHN------NDSNMNSIHNNDSNNNNNYKDCTHNPYICNESL 1206

Qy        517 ------NHLPQPNYA------------FTDTSDFSSLDDMNCHLNNNDKSIFLHKKNL-- 556
Db       1207 IIRNIMNRLIYQNIAQEDLFIPLEHDEFLVNKIMESYTDQECNYEDIIRSLDLN-KNVSY 1265

Qy        557 --RDNIKLKNRSSCD----NIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNI 610
Db       1266 IHNDQGKHLSLQMCAEEHITINNTNNDNTYV------EQIEMKELSKNKTKEKQSFKGTI 1319

Qy        611 NKMEDNTPKDI------------------------------LYESRNVSNMNGNVLLGL 639
Db       1320 RDMHEDSEEQMNKFITKKAKFFIEKKKGKMHECNDDIEYNNTQYDNIQYNNISCNYIKSQ 1379

Qy        640 N-KNTHHDISTKDENHNDNKINNGVI----------NIINN-SNVNSINN---SNMNSIN 684
Db       1380 NLENTHHQV-----NNDLSFIKNNVILPPKEYHSIFHFVNDYRNVVEIKNLMDKKKIFLN 1434

Qy        685 NSNMNSNSIYKSNYNS-----NQSISDVQIRYVNEMDTSNK------NNDNIFF---DAI 730
Db       1435 NSEKN---VVQSKYNRMSKNLKKKIEIINNIYRNEKKKLNRWKTKMDNDDNYWSSDDDSI 1491

Qy        731 -----------------------SCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRD 767
Db       1492 IAKKIIKIKNKEKRKYHPKEEKENFDRNNYKMITDNNNNDNNNNNNDNNNNDNNNNNNNS 1551

Qy        768 GTNNSMKLYAYNSH------------NLFQPDNN----KNTSNIQNINTNKNNQDGNVN 810
Db       1552 NNNN----YYYNLHDDVNNLGVTNYNTNIYPNDCNGIYEKETNN-NELTTNSNMCDKNND 1606

Qy        811 YSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDN--NIIQNRND------FEKKKKTNFY 862
Db       1607 FSDEFFN-NINENDLLYDNKYYRQIFKNVIGFVSVFEYVESYKQHYILFPYEIIKWTSFL 1665

Qy        863 NNNNIVIVNNNM-------------GNNNSPRMK-----------------YGLCGSHTS 892
Db       1666 LEYLTEIIPTNIFLHTKLSKKEKPTHQKNTGKMKIYIEEIKKWLFIKAINIYKYFSFKKS 1725

Qy        893 IDNMKNNEMKNNEMKDNEMKDNH------------------------------------- 915
Db       1726 IELIKKKDYFNYIIKNYDISHRYIIHDYSFINLKQLYLFIFFNIYKYFKYISTPGDAVGS 1785

Qy        916 ----------------------------------IKSNNNNSSSSSSNNNIYNNINDDD 941
Db       1786 ISAQSIGEPGTQMTLKTFHFAGVASMNVTLGVPRIKEIINASNSIQTPILNIPLEVNDNY 1845

Qy        942 TFQ-----------NDYC-HNDNTFTIRR-----KNNTN------INSNIYQNDDIIYT 977
Db       1846 NFALMMKSKLEKTTIRDICMYIKEDYTSRGVFLSVKFNEELIQKLFLNINAYNIKDIILK 1905

Qy        978 INSLNDYMSNTL-LHFKEKYTYPTLSTNEDIYNKEMEG-------------KHIR----- 1018
Db       1906 QSHINKIKINKIHINVINKYKLHISLKNDEFIFFQMESLKKGLLDLLIYGDKDIKRCIIK 1965

Qy       1019 -------------LDDQDKY---------------DDNDNNNVDNNNKNNVDNNV----- 1045
Db       1966 KEDIEVTDNEDEICDDMDEYYNVSQGTELYERKCNSKEENKNAIRVKKEEIDDNLEKEEN 2025

Qy       1046 ----------------------DNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNF 1082
Db       2026 IIYVSEKDSVNQLKSEKKKDINDDNNNNDDNNNNNDDDNKINDTIFNDDIDSDRNNLKEN 2085

Qy       1083 NNN---------EYLSYFQK----------------NVDTIINNCLNSLDISSMYDDTKE 1117
Db       2086 GSKLENVGEHIIERLSYKMKEKNVKKEHIKKEPNLINLDTINLDTLN-------FDE--- 2135

Qy       1118 ILNNILLSKYKAEKDNVIKKYIN-EDIKNMSLEEIDK-------TAQSIYEKRKVLLTKL 1169
Db       2136 ---------------------INVHNINNEKIEFYDEHLNICQGNKKHIQKKKKKTVYS 2174
```

```
Qy        1170 LLLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFY------------------AY 1211
               :|:   :::  :    |   |  :       || :|:  :  |                ||
Db        2175 ILVEGNSLNYVLGLEGVDFK------HIISNHVINVFQVLGIEAARITIINEIKKCVEAY 2228

Qy        1212 SRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKL 1271
               | :: || ::::|  |    : ::  : :| :       |:  | |:  |          |: |
Db        2229 S-IDIDIRHIMLLADIMAFTGDILGINRFGI----QKARQSTLMLASFEE-----TNEHL 2278

Qy        1272 -VQDNIKNSMDHNNI 1285
                 |    ||  : |||
Db        2279 FVSSFFKNVDEINNI 2293


RESULT 14
ATX1_PLAFA
ID   ATX1_PLAFA     STANDARD;      PRT;  1956 AA.
AC   Q04956;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Probable cation-transporting ATPase 1 (EC 3.6.3.-).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T9/96;
RX   MEDLINE=93132070; PubMed=8421054;
RA   Krishna S., Cowan G., Meade J.C., Wells R.A., Stringer J.R.,
RA   Robson K.J.;
RT   "A family of cation ATPase-like molecules from Plasmodium
RT   falciparum.";
RL   J. Cell Biol. 120:385-398(1993).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC       ATPases). Subfamily V.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X65738; CAA46646.1; -.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   TIGRFAMs; TIGR01494; ATPase_P-type; 7.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.
FT   DOMAIN        1     35       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     36     58       POTENTIAL.
FT   DOMAIN       59     61       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     62     80       POTENTIAL.
FT   DOMAIN       81    407       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    408    427       POTENTIAL.
FT   DOMAIN      428    440       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    441    462       POTENTIAL.
FT   DOMAIN      463   1818       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1819   1837       POTENTIAL.
FT   DOMAIN     1838   1845       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1846   1863       POTENTIAL.
FT   DOMAIN     1864   1881       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1882   1905       POTENTIAL.
FT   DOMAIN     1906   1928       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1929   1952       POTENTIAL.
FT   DOMAIN     1953   1956       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     496    496       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL      1760   1760       MAGNESIUM (BY SIMILARITY).
FT   METAL      1764   1764       MAGNESIUM (BY SIMILARITY).
FT   DOMAIN      246    251       POLY-ASN.
FT   DOMAIN      252    256       POLY-LYS.
FT   DOMAIN      937    941       POLY-ASN.
FT   DOMAIN     1344   1347       POLY-LYS.
FT   DOMAIN     1363   1372       POLY-ASN.
FT   DOMAIN     1680   1684       POLY-ASN.
SQ   SEQUENCE   1956 AA;  230285 MW;  AE708AAE99009335 CRC64;

  Query Match             7.9%;  Score 543;  DB 1;  Length 1956;
  Best Local Similarity   21.5%;  Pred. No. 1.2e-16;
  Matches  321;  Conservative  202;  Mismatches  412;  Indels  560;  Gaps   73;

Qy          22 KDSDIITVKNN--CTLYIDE-------PRYKVDMTKYIERHEFIVDKVFDDTVDNFTVYE 72
Db         540 KDDNMIHNKNSIISEYYIKDNMKNLHTSSKKKSITK--ERSNFLV-----QTIKSCLLKD 592

Qy          73 NTIKPLIIDLYENGCVC--------SCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAA 124
Db         593 HYIKEKKKEYYTNNTYCNDLHINDSTCSSY----------LLNSE--------------- 627

Qy         125 GDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRV 184
Db         628 -------------TKDAYCEYYNI--DHLCDINKKNMDI----NSKNELMGK------- 660

Qy         185 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAG-- 242
Db         661 YSKNELMGKTIKNELMGKYSKNELMGKYSKNEL-----MGKYSKNELMGKYSKNELMGKY 715

Qy         243 --SERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
Db         716 SKNELMGKTIKNQVGVDTNIYHMN------------CDNDYNY----------------- 746

Qy         301 GKSKSIMIANISPTISC--CEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
Db         747 ---------DYPCDYNCNNCNDTYHRLEYHN-----------INKDNSFN-----IPPEK 781

Qy         359 GSEMNASSIENVVIKSN-----------HLLSNNNNNKINRGKIN---------DKIERN 398
Db         782 NKSYNNISEH---IKINYPLLFEALACCHTLS-KVNNKIMGDVLEILMFNFTNCDMLINN 837

Qy         399 N--ILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKG-----------LINY----- 440
Db         838 NSFIIKEK--------------KKNCSYDFQKIDGDKNIGANDERCHLNNNLVSYNILKR 883

Qy         441 ---KSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINN 497
Db         884 FEFQSRLQRMSVIVKSTYGNNNDDNNDDDNNNDDDNNDDN------------------NN 925

Qy         498 NINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTDSFDFSSLDDMNCHLNNNDKSIFLHKKNLR 557
Db         926 DDNNNDDNNDDNNNNNYYYN-------------------------------IFC--KGSP 952

Qy         558 DNIK---LKNR--SSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKD-------NTF 605
Db         953 EKIKELCLKSKIPNNYDEILNKYTKQ--------GMRILSISYKRVKSKNINLLNVKRSF 1004

Qy         606 FKSNI---------NKMEDNTPKDILYESR-----------------------NVSNMN 632
Db        1005 VESNLHFLGFLIFTNNMKKNAP-DIIHNLQTSGCQCIMSTGDNVLTSIHVAKKCGIINSN 1063

Qy         633 ------GNVLLGLNKNT---------HHDISTKDENHNDNKINNGVINIINNSNVNSINN 677
Db        1064 VESIIIGDVIPVVGKNNKQKKKLVWYNHKNDTYLKGHDKTCIDNEFTSIQSQMNSDNICG 1123

Qy         678 SNM--NSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDN- 734
Db        1124 DNICGDNIYGDNICGDNIYGDNINGD-----------------NINGDNIYGDNINGDNI 1166

Qy         735 -----NMYPNITNNNNN-----------NNNNNNNNNNIDVENYNNRD-GTNNSMKLYA 777
Db        1167 NGDNINTYDNIYGDNYNLDYCPTEYHKCTYNNSILYRNNFLYKKENKKDKNYKNISTLYE 1226

Qy         778 YNSHNLFQPDN------NKNTSNIQNINTNK--------------------------- 802
```

```
                 : :::: | |         ||:  |:   : | |
Db        1227 HRTNDI-QFDKLCDILINKDPRNVNIVLTGKAFIFLKKKFYSFHLPYYEECKNIVHYIMK 1285

Qy         803 ----------NNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQK--DKNIHGCDNNIIQN- 849
                         || : |: |      |||: |     :   |||    :| ::     :: |
Db        1286 KKHKKIKNIINNHNSNLYY-----HYNIIDT--FVKRMNKEYMCFNKLLYKIQQKLLYNL 1338

Qy         850 -RNDFEKKK-KTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMK 907
                | ::|||   |:|: : : :: ||  |||                        ||| :
Db        1339 IHNLYKKKKYMNNYYDIDEVHLIGNNNNNNN------------------KNNSKEKKLPL 1380

Qy         908 DNEMKDNHIKSN--NNNSSSSSSNNNI----YNNINDDDTFQNDYCHNDNTFTIRRKNN 961
                |:||  ||: |  |:| : :: ::|||    |   ::  :  : | ||       |:| |
Db        1381 KNKMK--HIRKNESNDNITFNTYTSNNIHLSKYKYVHHKNYYYPDSCTN-----LRKKKN 1433

Qy         962 T---NINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDI-YNKEMEGKHI 1017
               :    |:   ||              |: : ||   : |    |   :|| |: :||
Db        1434 SLFYNLKKYIYYEK---------KKYLQHCLLK-HDNYKKVELPRIKDINYSYQMESIKT 1483

Qy        1018 R-----LDDQDKYDD---------NDNNNVDNN----NKNNVDNNVDNNNVDNNVDNNDK 1059
               |      | :| : :          ||:||| |      ||| : |     |:: |     :|
Db        1484 RNFIHSLSEQFAFSNLILSFYIIKNDDNNVYNKNYIYNKNYIYN---KNSICNKNYICNK 1540

Qy        1060 NNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEIL 1119
               | : | |                 || |  ||       | :||: |  :  |    :|        :
Db        1541 NYIYNKN-----------NIYNKNN----IYNKKNILTHAKSVL----LSGSSKKFLKFF 1581

Qy        1120 NNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDT 1179
               :||:          ||       ::||  :  :: |:    |   ::|       |||:  :
Db        1582 SNIIRRHKLKEKKN------KKNIKRYKMNHVNNTS-----KGHIILNMCTHGFKKDYSS 1630

Qy        1180 Q------INNETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESEN 1233
                          :||:   |: | |                 | |   :: ::    |  |  :
Db        1631 LKNKYRIVNNKRYMLKNDNV-----------------YDRHMYNLTDMYRGTQYGCSKK- 1672

Qy        1234 LRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNK------KLVQDNIKNSMDH 1282
                           :||         |:  || : || ||          |: |  | :: |
Db        1673 ------------KNK-------NIYMNNNNNILKNKINRFLEHLLVDKCKRNICH 1708
```

RESULT 15
CUT7_SCHPO
ID   CUT7_SCHPO   STANDARD;   PRT;   1085 AA.
AC   P24339;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein cut7.
GN   CUT7 OR SPAC25G10.07C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91015362; PubMed=2145514;
RA   Hagan I., Yanagida M.;
RT   "Novel potential mitotic motor protein encoded by the fission yeast
RT   cut7+ gene.";
RL   Nature 347:563-566(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: COULD BE A SPINDLE POLE BODY MOTOR. ON TRANSITION FROM
CC   G2 TO M PHASE OF THE CELL CYCLE, THE SPINDLE POLE BODY DUPLICATES;
CC   THE DAUGHTER POLE BODIES SEED MICROTUBULES WHICH INTERDIGITATE TO
CC   FORM A SHORT SPINDLE THAT ELONGATES TOSPAN THE NUCLEUS AT
CC   METAPHASE. MUTATIONS AT CUT7 BLOCK SPINDLE FORMATION.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC   SUBFAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X57513; CAA40738.1; -.
DR   EMBL; Z70691; CAA94636.1; -.
DR   PIR; T38378; T38378.
DR   HSSP; P17119; 3KAR.
DR   GeneDB_SPombe; SPAC25G10.07c; -.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW   Mitosis; Cell cycle; Phosphorylation; Repeat.
FT   DOMAIN   70   435   KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN   436   604   COILED COIL (POTENTIAL).
FT   DOMAIN   715   740   COILED COIL (POTENTIAL).
FT   DOMAIN   897   955   COILED COIL (POTENTIAL).
FT   NP_BIND   159   166   ATP (BY SIMILARITY).
FT   REPEAT   987   998
FT   REPEAT   999   1010
FT   MOD_RES   1011   1011   PHOSPHORYLATION (BY CDC2) (BY
FT   SIMILARITY).
FT   CONFLICT   34   61   SASNPRKRREPPTIDTGYPDRSDTNSPT ->
FT   LRAILGNDVSLLLLTL (IN REF. 1).
SQ   SEQUENCE   1085 AA; 122133 MW; 5669277875559D58 CRC64;

```
  Query Match             7.9%;  Score 538.5;  DB 1;  Length 1085;
  Best Local Similarity   20.7%;  Pred. No. 1e-16;
  Matches  235;  Conservative  223;  Mismatches  410;  Indels  265;  Gaps   43;

Qy           3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYID-------EPRYKVDMTKYIERHEF 55
                : | |||| |  :: |             |::| :| :          |   : | :  :   :
Db          71 TNINVVVRVRGRTDQE----------VRDNSSLAVSTSGAMGAELAIQSDPSSMLVTKTY 120

Qy          56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD 115
                 |||     |     ::||:: |::  :   ||   |:  ||||||:||||||  |        ||
Db         121 AFDKVFGPEADQLMLFENSVAPMLEQVL-NGYNCTIFAYGQTGTGKTYTMSGD----LSD 175

Qy         116 TPGIFQYAAG----DIFTFLNIYDKDNTK-GIFISFYEIYCGKLYDLL---QKRKMVAAL 167
```

```
Db        176 SDGILSEGAGLIPRALYQLFSSLDNSNQEYAVKCSYYELYNEEIRDLLVSEELRKPARVF 235

Qy        168 EN----------GKKEVVVKD----LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESS 213

Db        236 EDTSRRGNVVITGIEESYIKNAGDGLRLLR-----------EGSHRRQVAATKCNDLSS 283

Qy        214 RSHAILNIDLK-------------DINKNTS----LGKIAFIDLAGSERGADTVSQNKQ 255

Db        284 RSHSIFTITLHRKVSSGMTDETNSLTINNNSDDLLRASKLHMVDLAGSENIGRSGAENKR 343

Qy        256 TQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTI 315

Db        344 ARETGM-INQSLLTLGRVINALVEKAHHIPYRESKLTRLLQDSLGGKTKTSMIVTVSSTN 402

Qy        316 SCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISI------LDSKGSEMNASSIEN 369

Db        403 TNLEETISTLEYAARAKSIRNKP----QNNQLVFRKVLIKDLVLDIERLKNDLNATRKKN 458

Qy        370 VV------------IKSNHLLSNNNNNKINRGKINDKIERNNI--LKNKSFDKPREGFT 414

Db        459 GVYLAESTYKELMDRVQNKDLLCQEQARKLEVLDLNVKSSREQLQYVSKSNQEHKKEVEA 518

Qy        415 STFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNN----------- 462

Db        519 LQLQLVNSSTELESVKSENEK----LKNELVLEIEKRKKYETNEAKITTVATDLSQYYRE 574

Qy        463 ---------DNNDNNNDNNNNNNNDSSSMVNNMINHM--INNNINNNIN-----VNNNNN 506

Db        575 SKEYIASLYEKLDRTERNNKENENNFWNLKFNLLTMLRSFHGSFTDETNGYFTLLNDFNA 634

Qy        507 NNNNNNNSHNNHLPQPNYAFTD-TSDFSSLDD------MNCHLNNNDKSIFLHKKNLRDN 559

Db        635 SMEELLNTHSNQL---LISMTKITEHFQSLDEALQSARSSCAVPNS--SLDLIVSELKD- 688

Qy        560 IKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPK 619

Db        689 ----SKNSLLDALEHSLQDISMSSQKLGNGISSELIELQKDMKESY------------R 731

Qy        620 DILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSN 679

Db        732 QLVQELRSLYNL---------QHTHEE--------SQKELMYGVRN-----DIDALVKTC 769

Qy        680 MNSINNSNM---NSNSIYKSNYNSNQ-----SISDVQIRYVNEMDTSNKNNDNIFFDAIS 731

Db        770 TTSLNDADIILSDYISDQKSKFESKQQDLIANIGKIVSNFLQE------QNESLYTKA-- 821

Qy        732 CDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKN 791

Db        822 -------DILHSHLNDTNSNIRKANEI----------MNNRSEEFLRNA----------- 853

Qy        792 TSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCD------NN 845

Db        854 ASQAEIVGANKERIQKTVENGSQL----LDSKSKAIHSNSRSMYDHCLALAESQKQGVNL 909

Qy        846 IIQNRNDFEKKKKTNFYNNNN------IVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNN 899

Db        910 EVQTLDRLLQKVKEHSEDNTKEKHQQLLDLLESLVGNNDN------------LIDSIKTP 957

Qy        900 EMKNNEMKDNEMKDNHIKSNNNNS--SSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIR 957

Db        958 HTELQKITDHVLKGTTSLANHTNELLGLGDESLCNLETTIEDTSLVK-----LETTGDTP 1012

Qy        958 RKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNK 1010

Db       1013 SKRELPATPSWTRDSSLIKETTNLN--LDSDKKFVRETYTSSNQTNEPDVYDK 1063
```

```
RESULT 16
K125_ARATH
ID   K125_ARATH     STANDARD;      PRT;  1056 AA.
AC   P82266;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable 125 kDa kinesin-related protein.
GN   AT2G36200 OR F2H17.19.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
CC   -!- FUNCTION: RESPONSIBLE FOR MICROTUBULE TRANSLOCATION. MAY BE
CC       IMPORTANT FOR THE ORGANIZATION OF PHRAGMOPLAST-SPECIFIC ARRAYS OF
CC       MICROTUBULES (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: MICROTUBULE-ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC       SUBFAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AC006921; AAD21445.1; -.
DR   PIR; H84777; H84777.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    363       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      406    526       COILED COIL (POTENTIAL).
FT   NP_BIND      98    105       ATP (BY SIMILARITY).
SQ   SEQUENCE   1056 AA;  119267 MW;  7AD18A3FA9AB9211 CRC64;

  Query Match              7.6%;  Score 520.5;  DB 1;  Length 1056;
  Best Local Similarity   21.6%;  Pred. No. 6e-16;
  Matches  240;  Conservative  207;  Mismatches  424;  Indels  241;  Gaps   42;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMT-KYIERHEFIVDKVFDD 63

Db         13 VQVLLRCRPFSDDELRSNAPQVLT----CNDLQREVAVSQNIAGKHIDR-VFTFDKVFGP 67

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLG------SQPYGQSDT- 116

Db         68 SAQQKDLYDQAVVPIVNEVLE-GFNCTIFAYGQTGTGKTYTMEGECRRSKSAPCGGLPAE 126

Qy        117 PGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLL-----------QKRKMV 164

Db        127 AGVIPRAVKQIFDTLE--GQQAEYSVKVTFLELYNEEITDLLAPEDLSRVAAEEKQKKPL 184

Qy        165 AALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK 224

Db        185 PLMEDGKGGVLVRGLEEEIVTSANEIFTLLERGSSKRRTAETFLNKQSSRSHSLFSITI- 243
```

```
Qy        225 DINKNT-------SLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM 277
              | : |         ||:  :||||||  : : ::: : :  |  ||:|||  |    | |:
Db        244 HIKEATPEGEELIKCGKLNLVDLAGSENISRSGARDGRAREAG-EINKSLLTLGRVISAL 302

Qy        278 DSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNK 337
                 |:|:|||:|||::|||    |::|: :|| :|| : | |:||:|| |: | || :||
Db        303 VEHLGHVPYRDSKLTRLLRDSLGGRTKTCIIATVSPAVHCLEETLSTLDYAHRAKNIRNK 362

Qy        338 STCINEEDDTNT---------ERISILDSKGSEMNASSIEN-VVIKSNHLLSNNNNNKIN 387
                :|::   :|          ||:         :|: ||  :| | :         :  |:
Db        363 PE-VNQKMMKSTLIKDLYGEIERLK------AEVYASREKNGVYMPKERYYQEESERKVM 415

Qy        388 R-------GKINDKIERNNILKNKSFDKPRE------------GFTSTFGKYSSLNDID 427
                      |:|  :  ::   |::|   : ||            :  |      :| |  :
Db        416 AEQIEQMGGQIENYQKQLEELQDKYVGQVRECSDLTTKLDITEKNLSQTCKVLASTN--E 473

Qy        428 KIKKN----KKKGLI--------------------NYKSTLYNDNTINKKHNNNNNNNND 463
              ::||:     |:|  |                    | :     :::::::|     :  |
Db        474 ELKKSQYAMKEKDFIISEQKKSENVLVQQACILQSNLEKATKDNSSLHQKIGREDKLSAD 533

Qy        464 NNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPN 523
                   |      :|    : |:: ||: | :     |:::     |    :  :  :|| : :
Db        534 ----NRKVVDNYQVELSEQISNLFNRVASCLSQQNVHLQGVNKLSQSRLEAHNKAILEMK 589

Qy        524 YAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKN----LRDNIKLKNRSSC---------DN 570
                     : |  |        ||       : |||  |      |  :    |   |:|         |
Db        590 KKVKASRDLYS-----SHLEAVQNVVRLHKANANACLEEVSALTTSSACSIDEFLASGDE 644

Qy        571 IMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTP---KDILYESRN 627
                 :   : |  |  |   :: :|: : ::      |        :| :  |     : :: ||:|
Db        645 TTSSLFDELQSALSSHQGEMALFARELRQRFHTTM--EQTQEMSEYTSTFFQKLMEESKN 702

Qy        628 VSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSN 687
                                      |:    ||::||: :|:                |  ::::       : :|
Db        703 A-------------------ETRAAEANDSQINS-IIDFQKTYEAQSKSDTDKLIADLTN 742

Qy        688 MNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNN 747
              : |:                    ||   :|:    |  :|     ||  ||:|  :        : ||
Db        743 LVSS---------------HIRRQHELVDSRLHN---FKDAVSSNKTFLDEHVSAVNNL 783

Qy        748 NNNNNNNNNNIDVENYNN-RDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQD 806
                    :       ::  |  |:|  :  |    :      |  |    :   ::
Db        784 TKDAKRKWETFSMQAENEAREGADFSAAKHCRMELLLQQSVGHAESA------------ 830

Qy        807 GNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNI---HGCDNNIIQNRNDFE------KKK 857
                        | |   :  ::  | :: :|:  |  :          ||:|       ::| |         :
Db        831 --------FKHCKITHES-LKEMTSKQVTDVSSLVRSACDSN---EQHDAEVDSARTAAE 878

Qy        858 KTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKN-NEMKDNEMKDNHI 916
               |    |:::|:                          |:  |  :|   : ::: :|        :
Db        879 KDVTKNSDDII--------------------QQIERMSEDEKASVSKILEN------V 910

Qy        917 KSNNNNSSSSSSSNNNIYNNIND--DDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDI 974
              :|:        |                 ||     :|||     |   :  ||    | :
Db        911 RSHEKTLESFQQDQCCQARCIEDKAQETFQQQYMEYEPTGATPTKNEPEIPTKA------ 964

Qy        975 IYTINSLNDYMSNTLL-HFKEKYTYPTLSTNE 1005
                  || ||     ||:   |:|   :| : :| |
Db        965 --TIESLRAMPIETLVEEFRENNSYESFATKE 994
```

```
RESULT 17
K125_TOBAC
ID   K125_TOBAC     STANDARD;      PRT;  1006 AA.
AC   O23826;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   125 kDa kinesin-related protein.
GN   TKRP125.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A., SEQUENCE OF 149-163 AND 194-211, AND
RP   CHARACTERIZATION.
RC   STRAIN=cv. Bright Yellow 2;
RX   MEDLINE=97196959; PubMed=9044048;
RA   Asada T., Kuriyama R., Shibaoka H.;
RT   "TKRP125, a kinesin-related protein involved in the centrosome-
RT   independent organization of the cytokinetic apparatus in tobacco BY-2
RT   cells.";
RL   J. Cell Sci. 110:179-189(1997).
RN   [2]
RP   FUNCTION.
RC   STRAIN=cv. Bright Yellow 2;
RX   MEDLINE=95074298; PubMed=7983184;
RA   Asada T., Shibaoka H.;
RT   "Isolation of polypeptides with microtubule-translocating activity
RT   from phragmoplasts of tobacco BY-2 cells.";
RL   J. Cell Sci. 107:2249-2257(1994).
CC   -!- FUNCTION: RESPONSIBLE FOR MICROTUBULE TRANSLOCATION. MAY BE
CC       IMPORTANT FOR THE ORGANIZATION OF PHRAGMOPLAST-SPECIFIC ARRAYS OF
CC       MICROTUBULES.
CC   -!- SUBCELLULAR LOCATION: MICROTUBULE-ASSOCIATED.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING S PHASE. EXPRESSION
CC       INCREASES AS CELL MOVES FROM S PHASE TO M PHASE AND THEN DECREASES
CC       RAPIDLY AS CELL ENTERS THE G1 PHASE. EXPRESSION INCREASES AGAIN
CC       DURING THE G2 PHASE.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D83711; BAA23159.1; -.
DR   PIR; T02017; T02017.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    359       KINESIN-MOTOR.
FT   DOMAIN      371    522       COILED COIL (POTENTIAL).
FT   NP_BIND      95    102       ATP (BY SIMILARITY).
FT   CONFLICT    150    150       T -> Q (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    153    153       E -> F (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    161    161       D -> Q (IN REF. 1; AA SEQUENCE).
SQ   SEQUENCE   1006 AA;  113710 MW;  890C0E0F3504AA7D CRC64;

  Query Match             7.5%;  Score 513;  DB 1;  Length 1006;
  Best Local Similarity   20.2%;  Pred. No. 1.2e-15;
  Matches  234;  Conservative  209;  Mismatches  379;  Indels  338;  Gaps   46;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMT-KYIERHEFIVDKVFDD 63
              ::|::| || |  |  :       |   |   |     ::   |:|:|  |  ||||
Db         10 VQVLLRCRPFSNDELRNNAPQVVT----CNDYQREVAVSQNIAGKHIDR-IFTFDKVFGP 64

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLG------SQPYGQ-SDT 116
               :     :|:  | |:: :: | |  |: ||||||:|||||| |        | | |:
Db         65 SAQQRDLYDQAIVPIVNEVLE-GFNCTIFAYGQTGTGKTYTMEGECKRSKSGPNGELPQE 123

Qy        117 PGIFQYAAGDIFTFLNIYDKDNTK-GIFISFYEIYCGKLYDLL-----------QKRKMV 164
               |:   |   :|  |   :  | :  : ::| |:|  :: |||                :::| :
Db        124 AGVIPRAVKQVFDTL---ESQNAEYSVKVTFLELYNEEITDLLAPEDLKVALEDRQKKQL 180
```

```
Qy        165 AALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK 224
                 :|:||  |:|: |:    |:   |:   :   |   |:        | :|||||:: :| :
Db        181 PLMEDGKGGVLVRGLEEEIVTSANEIFTLLERGSAKRRTAETLLNKQSSRSHSLFSITI- 239

Qy        225 DINKNT-------SLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM 277
               | : |          ||:  :||||||  : : ::  : :  |   ||:||| |   | |:
Db        240 HIKEATPEGEELIKCGKLNLVDLAGSENISRSGAREGRAREAG-EINKSLLTLGRVINAL 298

Qy        278 DSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNK 337
                   |||:|||:||::|||   |::|: :|| :|| : | |:||:|| |: | || |||
Db        299 VEHLGHIPYRDSKLTRLLRDSLGGRTKTCIIATVSPAVHCLEETLSTLDYAHRAKNIKNK 358

Qy        338 STCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIER 397
                                                                 ::| |: :
Db        359 P-------------------------------------------------EVNQKMMK 367

Qy        398 NNILKNKSFDKPREGFTSTFGKYSSLNDIDKIK-----KNKKKGLINYKSTLYNDNTINK 452
               : ::|:             :|     :|:::|          :| |:  |  |   :
Db        368 STLIKD-----------LYG------EIERLKAEVYAAREKNGVYIPKERYYQE----- 404

Qy        453 KHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNN 512
                                           |:  :|  :  |  |       ::: |:
Db        405 ----------------------ENERKAMA-DQIEQM-------GVSIENHQKQFEELQ 433

Qy        513 NSHNNHLPQPNYAFTDTSDFS-SLDDMNCHLNNNDKSIFLHKKNLRDN-IKLKNRSSCDN 570
               : |:: : |        || :  ||     ||    | :   :: || :    || |    :
Db        434 SRHDSQVQQ-------CSDLTCKLDVTQKQLNQTSKLLAYTEEQLRQSQYTLKERDFIIS 486

Qy        571 IMNKKKNNL-HLA-------RHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDIL 622
                 | :| | |            |:     ::|     :::| :|  :| :|  :      | :
Db        487 EQKKAENALAHQACVLRADLEKSIQENASLFQKIAREDKLSTDNRSLVNNFQAELAKQL- 545

Qy        623 YESRNVSNMNGNVLLGLNKNTHHDISTKDENHND-NKINNGVINIINNSNVNSINNSNMN 681
                    ::: :  : |   |   : | :  |::: |
Db        546 ------GSLSSTLATSVCRQTEHLQCVEKFCHNFLDSHDKAVLDLKRKIN---------- 589

Qy        682 SINNSNMNSNSIYKSNYNSNQSI-------SDVQIRYVNEMDTSNKNNDNIFFDAISCD- 733
                    :| ::| |:: : |::         |:   :   |: : :||   :    | || : :
Db        590 -------SSMALYISHFEAMQNVVRLHKATSNATLEEVSTLASSNSISTKEFLDAEAVEA 642

Qy        734 NNMYPNITNNNNNNN--------------NNNNNNNNID--VENYNNRDGTNNSMKLYA 777
               |:|:  : :  : :                  |::  :  ||   :: : ::   : | :|
Db        643 NSMFDELQSTLSTHQGEMAHFARELRQRFNDSTEHLTNISAIIQRFFDK-LLDESKRL-- 699

Qy        778 YNSHNLFQPDNNKNTSNIQNINTNK--NNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQK 835
                          |: : :  | ||          :        | :        :   |   |: : :||
Db        700 ----------EKHATTVDEIQTNSIAEFEKAYEEQSKSDAEKLIADVTSLVSNHMRRQK 748

Qy        836 D-----------------------KNIHGCDNNIIQNRNDF------EKKKKTNF---- 861
              :                          :: |   :  :  ||          | |: :|
Db        749 ELVGARLVDLRETVSGNRTFLDGHVSSMEGITTDAKRKWQDFYMQAEGETKENADFSAAK 808

Qy        862 ---------------------YNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNM-KN 898
                                     : : :|   .|:|||  :   |          |: : |:  |
Db        809 HCRMESLMQKCVSTAETALKRWQSTHELV---NDMGNQHVLTM-------HSVVRNICDN 858

Qy        899 NEM------KNNEMKDNEMKDNH---IKSNNNSS----------SSSSSNNNIYNNIND 939
               ||         |  : ::| |     ||| :: |                ::|::: : :
Db        859 NEQHVTDFDSTRESAEEDVKRNSEDIIKSIDSLSGEERGSISGVLDTTSAHSETLDVLKK 918

Qy        940 D-------------DTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMS 986
               |              :|||  |    :  |       ::  :: | :            || ||
Db        919 DHCMQSTSIEQIALETFQQKYMDYEPTGATPIRSEPDVPSKV--------TIESLRAMPM 970

Qy        987 NTLL-HFKEKYTYPTLSTNE 1005
                 ||  |:|   :: :     |
Db        971 EVLLEEFRENNSFESFQVKE 990

RESULT 18
KINH_LOLPE
ID   KINH_LOLPE     STANDARD;      PRT;   967 AA.
AC   P21613;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Kinesin heavy chain.
OS   Loligo pealeii (Longfin squid).
OC   Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC   Decapodiformes; Loliginidae; Loligo.
OX   NCBI_TaxID=6621;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90153980; PubMed=2137456;
RA   Kosik K.S., Orecchio L.D., Schnapp B., Inouye H., Neve R.L.;
RT   "The primary structure and analysis of the squid kinesin heavy
RT   chain.";
RL   J. Biol. Chem. 265:3278-3283(1990).
CC   -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC       PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC   -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC       CHAINS.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC       TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC       KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC       ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC       DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC       INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC       VESICLES AND MEMBRANOUS ORGANELLES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       SUBFAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J05258; AAA29990.1; -.
DR   PIR; A35075; A35075.
DR   HSSP; P33176; 1BG2.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    385       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      392    861       COILED COIL.
FT   DOMAIN      862    967       GLOBULAR.
FT   DOMAIN      173    314       MICROTUBULE-BINDING.
FT   NP_BIND      85     92       ATP (BY SIMILARITY).
SQ   SEQUENCE   967 AA;  109407 MW;  CC40353FA3F33E35 CRC64;

  Query Match             7.3%;  Score 500.5;  DB 1;  Length 967;
  Best Local Similarity   22.1%;  Pred. No. 4e-15;
  Matches  217;  Conservative  189;  Mismatches  403;  Indels  175;  Gaps   35;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 64
              |||: | |||:| |::       |:                 :: | :  |  |: |||
Db          9 IKVICRVRPLNEAEERAGSKFIL-------------KFPTDDSISIAGKVFVFDKVLKPN 55

Qy         65 VDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTP---GIFQ 121
              |    ||   ||:: |:  :||  : ||||| ||||:||         |   |   ||
Db         56 VSQEYVYNVGAKPIVADVL-SGCNGTIFAYGQTSSGKTHTM-----EGVLDKPSMHGIIP 109

Qy        122 YAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKI 181
                  ||| ::   |::   | ||:||||  |: |||   |    |:   |   |
Db        110 RIVQDIFNYIYGMDENLEFHIKISYYEIYLDKIRDLLDVTKTNLAVHEDKNRVPFVKGAT 169

Qy        182 LRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTS---LGKIAF 237
                | ::  | ::::|| |   | : | : |: |||||::  |::|  |  |     ||:
```

```
Db        170 ERFVSSPEEVMEVIDEGKNNRHVAVTNMNEHSSRSHSVFLINVKQENVETQKKLSGKLYL 229

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM-DSDKNHIPFRDSELTKVLR 296
              :||||||: : | ::        :  |||:|| ||    | |: | :|:|:|:|||:||::|:
Db        230 VDLAGSEKVSKTGAEG-AVLDEAKNINKSLSALGNVISALADGNKSHVPYRDSKLTRILQ 288

Qy        297 DIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEE---------DDT 347
               :   | ::: |:    ||        :| :|| :   | |  ||     :|||          :
Db        289 ESLGGNARTTMVICCSPASYNESETKSTLLFGQRAKTIKN-VVSVNEELTADEWKRRYEK 347

Qy        348 NTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKI---ERNNILKNK 404
                ||:: | :  ::: |                    :: | :    :     |: :: ::
Db        348 EKERVTKLKATMAKLEA--------------------ELQRWRTGQAVSVEEQVDLKEDV 387

Qy        405 SFDKPREGFTS-TFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKK---------- 453
                : |    ||   |  :|:|: |: :  :::  | |:       |: || :
Db        388 PAESPATSTTSLAGGLIASMNEGDRTQLEEER-LKLYQQLDDKDDEINNQSQLIEKLKEQ 446

Qy        454 ---------HNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNN 504
                         :   :   |    :  :    :| :   |      |     :   :   |      :   | :
Db        447 MMEQEDLIAQSRRDYENLQQDMSRIQADNESAKDEVKEVLQALEELAMNYDQKSQEVEDK 506

Qy        505 NNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNI--KL 562
              |  |  | :    |  |                    |:|: :   |:      |    |:| : | :    |
Db        507 NKENENLSEELNQKL------------STLNSLQNELDQLKDSSMHHRKRVTDMMINLL 553

Qy        563 KNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDIL 622
              |:          |:      |     : :  ||                :  ::  |   :   |:||:
Db        554 KDLGDIGTIVG---GNAAETKPTAGS------GEKIEEEFTVARLYISKMKS------- 596

Qy        623 YESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNS 682
               | : :  :  |  :   :||    | ||  |    |   |   ::    :  :  : |:  :  :|
Db        597 -EVKTLVSRNNQL-----ENTQQDNFKKIETHEKDLSNCKLLIQQHEAKMASLQEAIKDS 650

Qy        683 INNSNMNSNSIYKSNYNSNQSISDVQIRYV----NEMDTSNKNNDNIFFDAISCDNNMYP 738
                |   |   :::    |     :   :   |:              | :||   :        :   :         |:
Db        651 ENKKRMLEDNVDSLNEEYAKLKAQEQMHLAALSEREKETSQASETR---EVLEKQMEMHR 707

Qy        739 NITNNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSM---KLYAYNSHNLFQPDNNKNTSNI 795
                      :        :   :   :           |:|   :   |      |:    ||   |     ::   :   :   :   :   :
Db        708 E-QHQKQLQSLRDEISEKQATVDNLKD-DNQRLSLALEKLQA--DYDKLKQEEVEKAAKL 763

Qy        796 QNINTNKNNQ----------DGNVNYSMNFCHYNLNDKNYLIDLNNK--------EQKDK 837
               :::     : :                :     |     :        | ||    |   :: || ||           |::|:
Db        764 ADLSLQIDRREQAKQDLKGLEETVAKELQTLH-NLR-KLFVQDLQNKVKKSCSKTEEEDE 821

Qy        838 NIHGCDNNIIQNRNDFEKKKKTNFYNNN--NIVIVNNNMGNNNS------PRMKYGLCGS 889
               :     |              :|:|: :|   ||     :    |:  :   :   :|:          |::|
Db        822 DTGG----------NAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRL--- 868

Qy        890 HTSIDNMKNNEMKNNEMKDNEMKD 913
                ::: :|: |      : |:   :|
Db        869 RATMERVKSLESALKDAKEGAMRD 892
```

```
RESULT 19
BIMC_EMENI
ID   BIMC_EMENI     STANDARD;      PRT;  1184 AA.
AC   P17120;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Kinesin-like protein BIMC.
GN   BIMC.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=162425;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90199865; PubMed=2138511;
RA   Enos A.P., Morris N.R.;
RT   "Mutation of a gene that encodes a kinesin-like protein blocks
RT   nuclear division in A. nidulans.";
RL   Cell 60:1019-1027(1990).
CC   -!- FUNCTION: IMPORTANT ROLE IN MITOTIC DIVIDING CELLS. MICROTUBULE
CC       MOTOR REQUIRED FOR SPINDLE BODY SEPARATION.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M32075; AAA33298.1; -.
DR   PIR; A34795; A34795.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW   Mitosis; Cell cycle; Phosphorylation.
FT   DOMAIN       79    485       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      489    900       COILED COIL (POTENTIAL).
FT   DOMAIN      901   1184       GLOBULAR (POTENTIAL).
FT   NP_BIND     167    174       ATP (BY SIMILARITY).
FT   MOD_RES    1006   1006       PHOSPHORYLATION (BY CDC2) (BY
FT                                SIMILARITY).
SQ   SEQUENCE   1184 AA;  131630 MW;  CEA015EC8F980E4F CRC64;

  Query Match             7.3%;  Score 500;  DB 1;  Length 1184;
  Best Local Similarity   21.1%;  Pred. No. 5.3e-15;
  Matches  253;  Conservative  194;  Mismatches  415;  Indels  338;  Gaps   48;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDM-TKYIERHEFIVDKV 60
              :: | |||| | :| | |:     :: :      :     :: |    :     : |||
Db         79 DTSIHVVVRCRGRNEREVKENSGVVLQTEG-----VKGKTVELSMGPNAVSNKTYTFDKV 133

Qy         61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120
               |     | ||||: : |:: ::     |:  |||||||:||||||  |          :|| ||
Db        134 FSAAADQITVYEDVVLPIVTEMLA-GYNCTIFAYGQTGTGKTYTMSGD----MTDTLGIL 188

Qy        121 QYAAGDI-----FTFLNIYDKDNTKGIFISFYEIYCGKLYDLL-------------QKR 161
              ·  || |        :    |  | ::|   || |:|  :| |||                    ||:
Db        189 SDNAGIIPRVLYSLFAKLADTEST--VKCSFIELYNEELRDLLSAEENPKLKIYDNEQKK 246

Qy        162 KMVAALENGKKEVVVKD----LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHA 217
               ::  |   | :|   :        :|:|:                  |    |::        || |||||
Db        247 GHMSTLVQGMEETYIDSATAGIKLLQ------------QGSHKRQVAATKCNDLSSRSHT 294

Qy        218 ILNIDLKDINKNTSL-------GKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLAL 270
               :  | : :| : |              ||:  :||||||     : ::||:  |:   ||:||| |
Db        295 VFTITV-NIKRTTESGEEYVCPGKLNLVDLAGSENIGRSGAENKRA-TEAGLINKSLLTL 352

Qy        271 KECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSR 330
                 | |:      |||:|:|:||::|:|  |::|: :|| :||  |  |  |:|::|| |: |
Db        353 GRVINALVDKSQHIPYRESKLTRLLQDSLGGRTKTCIIATMSPARSNLEETISTLDYAFR 412

Qy        331 VKNFKNKSTCIN-------------EEDDTNTERISILDSKGSEMNASSIENVVI--KS 374
              || :||   ||                    :       |     |:    |     |  |  : :  |
Db        413 AKNIRNKPQ-INSTMPKMTLLREFTAEIEKLKAELIATRHRNGVYMSVESYEEMKMENES 471

Qy        375 NHLLSNNNNNKIN--RGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDI----DK 428
                ::|     ||        :    |::    | :| |:  ::      |      | ||:    |
Db        472 RRIISEEQRAKIESMESSLRHKVQELLTLTSK-FNDLKKDNDDTLAALCSTNDVLQQTDI 530

Qy        429 IKKNKK----------------------KGLIN-YKSTLYNDNTINKKHNNNNNNNND 463
               : :| :                              ||||:     |: : |::   |        |
Db        531 VLQNTRAQLEEEEMLRCAHEETEHQLQDVGKGLISTLGQTVEDINSLQSK--------LD 582
```

```
Qy        464 NNDNNNDNNNNNNNNDSSSMVNNMINHMINNNI------------NNNINVN-------NN 504
                  :    |       ||: |:: :    |:  :              :: ||          :|
Db        583 RKAELDATNAELWRASSTEVSD-VTKRIDQRVEAFQTRHAKLLETTSVKVNEFIATEISN 641

Qy        505 NNNNNNNNNSHNNHLPQP-NYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLK 563
                  :: : :|  |      | |  :||  |: :|||           ::       |:||: :| |
Db        642 IERTRSDLSEYNRSLDAACNNAKAETS--SAHEDMN--------NVLEEIKDLREEVKSK 691

Qy        564 NRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILY 623
                      : :  | |  |    :   ::              : |    | ::    :|  ||:
Db        692 --------VGEGLNGLSAAAARISEEVI---------GEFTQLHSQLHTSFNNLGKDL-- 732

Qy        624 ESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSI 683
                                |:     ::|    :    |:||          : :  |  |   |: :
Db        733 ----------------KSIFETMAT-HLSEQKNEINR------LRAELQSSNRQNIETT 768

Qy        684 NNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDA--------ISCDNN 735
               : :: :      :  : : ::  :: :  :    ::           |          || :
Db        769 HKASAHLAQAIEEEHVAAEAEREILMSQIKALVEESRQKQFARLRAKIDGVRTEISASGD 828

Qy        736 MYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNI 795
               |    |                  :: ||   :                        |:
Db        829 MLEQAT---------TQHDRQIDEWVF-------------------------KSEQFA 852

Qy        796 QNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEK 855
               :::| :|:                           ::  | | |       |: |
Db        853 KDVNASKD-----------------------EIRTKLQNDWEAFDQRNSTI-------- 880

Qy        856 KKKTNFYNNNNIVIVN---NNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMK 912
               :| |    :   : ||:    ::||                  ::|:         |         :
Db        881 RKATESVHKETVRIVDVQVDDMGRQME------------ALDDFVAKARSQN----GRYR 924

Qy        913 DNHIKSNNNNSS----SSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIR--------RKN 960
               | ||  :  :   ::       | ||       :  |        :  ||  :    |:   |           ||
Db        925 DAHIATLDTIATGVRDSYSSIEGRVENLTGRMNQFQQEATHHHATLEESIAPLSNDVRKP 984

Qy        961 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTY----PTLSTNEDIYN-----KE 1011
               |:::|: :||          || :|::  :    | || |      |:  ::| : :       ||
Db        985 LTDLSSS-FQN-------RSLEEYVATGVTPKKRKYDYISVLPSTESHEVLKSRLRTTKE 1036

Qy       1012 MEGKHIRLDDQDK--------------YDDND-----------NNNVDNNNKNNVDNNV 1045
               ||       |||                       |:|  :           | |   |       || ||
Db       1037 MEVLPFNSDDQLSGPSSSPGGSPSKGFVYNDVEDEVGTHAPTVTNVNPSNTGLREVDANV 1096
```

```
RESULT 20
CENE_HUMAN
ID   CENE_HUMAN     STANDARD;      PRT;  2663 AA.
AC   Q02224;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Centromeric protein E (CENP-E protein).
GN   CENPE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93024922; PubMed=1406971;
RA   Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT   "CENP-E is a putative kinetochore motor that accumulates just before
RT   mitosis.";
RL   Nature 359:536-539(1992).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=95196755; PubMed=7889940;
RA   Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT   "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT   microtubule motor.";
RL   EMBO J. 14:918-926(1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98437347; PubMed=9763420;
RA   Chan G.K.T., Schaar B.T., Yen T.J.;
RT   "Characterization of the kinetochore binding domain of CENP-E reveals
RT   interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL   J. Cell Biol. 143:49-63(1998).
CC   -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC       KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC       OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC       AND/OR SPINDLE ELONGATION.
CC   -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC   -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC       CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC       QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z15005; CAA78727.1; -.
DR   PIR; S28261; S28261.
DR   HSSP; P17119; 3KAR.
DR   Genew; HGNC:1856; CENPE.
DR   GK; Q02224; -.
DR   MIM; 117143; -.
DR   GO; GO:0005699; C:kinetochore; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0008350; F:kinetochore motor activity; TAS.
DR   GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR   GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR   GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW   Cell cycle; Centromere.
FT   DOMAIN        1    335       KINESIN-MOTOR.
FT   DOMAIN      336   2471       COILED COIL (POTENTIAL).
FT   DOMAIN     2472   2663       GLOBULAR (POTENTIAL).
FT   NP_BIND      86     93       ATP (BY SIMILARITY).
SQ   SEQUENCE   2663 AA;  312087 MW;  CEFC13880C8C8CB8 CRC64;

  Query Match             7.1%;  Score 482.5;  DB 1;  Length 2663;
  Best Local Similarity   18.7%;  Pred. No. 7.3e-14;
  Matches  306;  Conservative  258;  Mismatches  612;  Indels  457;  Gaps   58;

Qy          5 IKVVVRKRPLSELEKK---------KKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEF 55
                : | || |||:  |:               | |:::|             |:|| :|      |
Db          7 VAVCVRVRPLNSREESLGETAQVYWKTDNNVI--------------YQVDGSK-----SF 47

Qy         56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD 115
                 |:||             |||     | |||      |     :  ||||| ||||||:||:      |
Db         48 NFDRVFHGNETTKNVYEEIAAP-IIDSAIQGYNGTIFAYGQTASGKTYTMMGSE-----D 101

Qy        116 TPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLL---QKRKMVAALENGKK 172
                 |:   |  |||  :   :   |      : :|: |||     : |||    || |  :    |:  :
Db        102 HLGVIPRAIHDIFQKIKKF-PDREFLLRVSYMEIYNETITDLLCGTQKMKPLIIREDVNR 160

Qy        173 EVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINK---- 228
                | | ||     | | |   :   :    |     |     |      | |||||  |   : |:   |
Db        161 NVYVADLTEEVVYTSEMALKWITKGEKSRHYGETKMNQRSSRSHTIFRMILESREKGEPS 220

Qy        229 ----NTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDK--N 282
                  :   :   :|||||||  | |  :     : : :| ||||||   | : |: :    :
```

```
Db        221 NCEGSVKVSHLNLVDLAGSERAAQTGAAGVRLK-EGCNINRSLFILGQVIKKLSDGQVGG 279

Qy        283 HIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCIN 342

Db        280 FINYRDSKLTRILQNSLGGNPKTRIICTITPV--SFDETLTALQFASTAKYMKN-TPYVN 336

Qy        343 E--EDDTNTERI--SILDSK------GSEMNASSIEN----VVIKSNHLLSNNNNNKINR 388

Db        337 EVSTDEALLKRYRKEIMDLKKQLEEVSLETRAQAMEKDQLAQLLEEKDLLQKVQNEKIEN 396

Qy        389 ----------------------------GKINDKIERN--------------------N 399

Db        397 LTRMLVTSSSLTLQQELKAKRKRRVTWCLGKINKMKNSNYADQFNIPTNITTKTHKLSIN 456

Qy        400 ILK--NKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNN- 456

Db        457 LLREIDESVCSESDVFSNT------LDTLSEIEWNPATKLLNQENIESELNSLRADYDNL 510

Qy        457 ---------------NNNNNNDNNDNNNDNNNNNNNDSSSMVN------NMINHM---- 490

Db        511 VLDYEQLRTEKEEMELKLKEKNDLDEFEALERKTKKDQEMQLIHEISNLKNLVKHREVYN 570

Qy        491 --INNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDD---------- 537

Db        571 QDLENELSSKVELLREKEDQIKKLQEYIDSQKLENIKMDLSYSLESIEDPKQMKQTLFDA 630

Qy        538 MNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDP 597

Db        631 ETVALDAKRESAFLRSENLELKEKMKELAT---TYKQMENDIQLYQSQLEAKKKM-QVDL 686

Qy        598 QKNKDNTFFKSNINKM----EDNTPKDILYE------------------SRNVSNMNGNV 635

Db        687 EKELQSAF--NEITKLTSLIDGKVPKDLLCNLELEGKITDLQKELNKEVEENEALREEVI 744

Qy        636 LLGLNKNTHHDI------------------STKDE-----NHNDNKINNGVINIINNSNV 672

Db        745 LLSELKSLPSEVERLRKEIQDKSEELHIITSEKDKLFSEVVHKESRVQGLLEEIGKTKDD 804

Qy        673 NSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRY---VNEMDTSNKNNDNIFFDA 729

Db        805 LATTQSNYKSTDQEFQNFKTLH----------MDFEQKYKMVLEENERMNQEIVNLSKEA 854

Qy        730 ISCDNNM--------YPNITNNNNNNNNNNNNNNIDVENYNNRDG------------- 768

Db        855 QKFDSSLGALKTELSYKTQELQEKTREVQERLNEMEQLKEQLENRDSPLQTVEREKTLIT 914

Qy        769 -----TNNSMKLYAYNSHNL------FQPDNNKNTSNIQ---NINTNKNNQDGNVNYSMN 814

Db        915 EKLQQTLEEVKTLTQEKDDLKQLQESLQIERDQLKSDIHDTVNMNIDTQEQLRNALESLK 974

Qy        815 FCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNM 874

Db        975 QHQETINTLKSKI----SEEVSRNLH-MEENTGETKDEFQQKM----------------- 1012

Qy        875 GNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSNNNIY 934

Db       1013 ----------------VGIDKKQDLEAKNTQTLTADVKDNEIIEQQRKIFSLIQEKNELQ 1056

Qy        935 NNI----------------NDDDTFQN--------DYCHNDNTFTIRRKNNTNINSNIYQ 970

Db       1057 QMLESVIAEKEQLKTDLKENIEMTIENQEELRLLGDELKKQQEIVAQEKNHA-----IKK 1111

Qy        971 NDDIIYTINSLNDY-------------MSNTLLHFKEKYTYPTLSTNE------DIYNKE 1011

Db       1112 EGELSRTCDRLAEVEEKLKEKSQQLQEKQQQLLNVQEEMSEMQKKINEIENLKNELKNKE 1171

Qy       1012 MEGKHI---RLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDN---- 1064

Db       1172 LTLEHMETERLELAQKL---------NENYEEVKSITKERKVLKELQKSFETERDHLRGY 1222

Qy       1065 -NNVD----NDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMY------- 1112

Db       1223 IREIEATGLQTKEELKIAHIHLKEHQETIDELRRSVSEKTAQIINTQDLEKSHTKLQEEI 1282

Qy       1113 ----------------DDTKEILNNI-LLSKYKAEKDNVIKKYI--------------NE 1141

Db       1283 PVLHEEQELLPNVKKVSETQETMNELELLTEQSTTKDSTTLARIEMERLRLNEKFQESQE 1342

Qy       1142 DIKNMS-------------------LEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQI 1181

Db       1343 EIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQESQSKQEQSLNMKEKDNETTKI 1402

Qy       1182 NNETSD--------LRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESEN 1233

Db       1403 VSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDEMKSVAKEKDDLQRLQEVLQ--SESDQ 1460

Qy       1234 LRLLYQFLVVEY----------------QNKSANSVLLNVSSNNGDIILLNKKL--VQDN 1275

Db       1461 LKENIKEIVAKHLETEEELKVAHCCLKEQEETINELRVNLSEKETEISTIQKQLEAINDK 1520

Qy       1276 IKNSMDHNNIHKK 1288

Db       1521 LQNKI--QEIYEK 1531
```

RESULT 21
KL68_DROME
ID KL68_DROME STANDARD; PRT; 784 AA.
AC P46867; Q9VTN8;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kinesin-like protein KLP68D.
GN KLP68D OR KLP5 OR CG7293.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95050960; PubMed=7525600;
RA Pesavento P.A., Stewart R.J., Goldstein L.S.B.;
RT "Characterization of the KLP68D kinesin-like protein in Drosophila:
RT possible roles in axonal transport.";
RL J. Cell Biol. 127:1041-1048(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

```
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE OF 220-342 FROM N.A.
RX   MEDLINE=92020874; PubMed=1924306;
RA   Stewart R.J., Pesavento P.A., Woerpel D.N., Goldstein L.S.B.;
RT   "Identification and partial characterization of six members of the
RT   kinesin superfamily in Drosophila.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:8470-8474(1991).
CC   -!- FUNCTION: PLUS-END DIRECTED MICROTUBULE MOTOR THAT MAY BE USED FOR
CC       ANTEROGRADE AXONAL TRANSPORT AND COULD CONCEIVABLY MOVE CARGOES IN
CC       FLY NEURONS DIFFERENT THAN THOSE MOVED BY KINESIN HEAVY CHAIN OR
CC       OTHER PLUS-END DIRECTED MOTORS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN THE CENTRAL NERVOUS
CC       SYSTEM AND IN A SUBSET OF THE PERIPHERAL NERVOUS SYSTEM DURING
CC       EMBRYOGENESIS.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       II SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U15974; AAA69929.1; -.
DR   EMBL; M74431; AAA28658.1; -.
DR   EMBL; AE003543; AAF50008.1; -.
DR   PIR; A55236; A55236.
DR   HSSP; P17119; 3KAR.
DR   FlyBase; FBgn0004381; Klp68D.
DR   GO; GO:0003774; F:motor activity; IDA.
DR   GO; GO:0008089; P:anterograde axon cargo transport; IEP.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN       16    275       KINESIN-MOTOR.
FT   DOMAIN      351    385       COILED COIL (POTENTIAL).
FT   DOMAIN      426    582       COILED COIL (POTENTIAL).
FT   NP_BIND     106    113       ATP (POTENTIAL).
FT   CONFLICT    220    221       SS -> TC (IN REF. 3).
FT   CONFLICT    338    342       GSRAK -> VRGQV (IN REF. 3).
FT   CONFLICT    338    338       G -> A (IN REF. 2).
SQ   SEQUENCE   784 AA;  88193 MW;  94BB9BADF072DFC0 CRC64;

  Query Match              7.0%;  Score 477;  DB 1;  Length 784;
  Best Local Similarity   22.4%;  Pred. No. 3.4e-14;
  Matches  185;  Conservative  130;  Mismatches  291;  Indels  220;  Gaps   27;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYI-----ERHEFI 56
                |  ::||||  ||:|   |: ::  :::  |   |        |  |::     :    :|  |
Db         17 NECVQVVVRCRPMSNRERSERSPEVVNVYPN--------RGVVELQNVVDGNKEQRKVFT 68

Qy         57 VDKVFDDTVDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGSQPYGQS 114
                 |  :|  :     |:|    :  ||:   :  |   |||:    |||||||:|||:|| |  :  |
Db         69 YDAAYDASATQTTLYHEVVFPLVSSVLEGFNGCI---FAYGQTGTGKTFTMEGVR--GND 123

Qy        115 DTPGIFQYAAGDIFTFLNIYDKDNTKG----IFISFYEIYCGKLYDLLQKRKMVAALENG 170
              :  ||        |:  :|       |:       : :|: |||  :| |||:       :
Db        124 ELMGIIPRTFEQIWLHIN-----RTENFQFLVDVSYLEIYMEELRDLLKPNSKHLEVRER 178

Qy        171 KKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAI--LNIDLKDINK 228
                  | |  :|  :    : |::|   |    |     | :|   :  |:  ||||||   :  |:: |
Db        179 GSGVYVPNLHAINCKSVEDMIKVMQVGNKNRTVGFTNMNEHSSRSHAIFMIKIEMCDTET 238

Qy        229 NT-SLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFR 287
              ||  :||:  ||||||| : |  :  :: : : :  || :| :|     | |:       |:|:|
Db        239 NTIKVGKLNLIDLAGSERQSKTGASAERLK-EASKINLALSSLGNVISALAESSPHVPYR 297

Qy        288 DSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDT 347
              ||:||::|:|     | ||:||||||  |:        :||  ||||  ||  |: :|:    |   ||
Db        298 DSKLTRLLQDSLGGNSKTIMIANIGPSNYNYNETLTTLRYGSRAKSIQNQP--IKNEDPQ 355

Qy        348 NT---------ERI----------------------------SILDSKGSEMNASSIEN 369
              :           ||:                            ::          :  |:||
Db        356 DAKLKEYQEEIERLKRLIGPQQQQRSEKQVTAKKQRVKKPKKETVTKEMSDSLQVSTIEQ 415

Qy        370 VVIKSNH----LLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE-------GFTSTFG 418
               |    :         ::  |:      |  |:::||   :  :|         |   |         |    :
Db        416 PVEDDSDPEGAESESDKENEAEVAKSNEELERERVENSKLAAKLAELEGQLVRGGKNLLD 475

Qy        419 KYSSLN-DIDK----IKKNKKKGL-INYKSTLYNDNTINKKHNNNNNNNNDN-------- 464
                ||    :::|      | :  ||: :  |    :    |   :  |:   :    |   :
Db        476 TYSERQIELEKKLVEIAERKKREIEIQQLELQEETTLEIRERNVSLEQEVELKKRKLSK 535

Qy        465 --------NDNNNDNNNNNNND--SSSMVNN----------MINHMINNNINNNINVNN 503
                          ||  :::|  |        |   |          :|::  :     :     :
Db        536 CYAKYLALQQELNDCKSDHNQDLRELEMAQNELVKELKRQLLIIDNFVPIEVKQRLYTQA 595

Qy        504 NNNNNNNNNNSHNNHLPQPNYAFTDTDSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLK 563
                 :          :   ||  |
Db        596 KYDEEQEEWKFSSMSLPTP------------------------------------------ 614

Qy        564 NRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILY 623
                   |   : |:   |                  ||: :  :   :   :  :  :|:|  :   :
Db        615 --PGGDGKFSSKRPVSH----------------PQRRRPTSEYA--LQEAKSNSPSSLRF 654

Qy        624 ESRNVSNMN-------------------GNVLLGLNKNTHHDISTKDENHNDNKINNGV 663
              :|  |:  |                   ||    :    ||    |    :  |  :  :  :
Db        655 KSENIVNYELEMPCRTTQEYRTPKVSASLQAVLAQAMQTGDDIDIVDS--HTNSLRSRL 712

Qy        664 INIINNS-----------------NVNSINNSNMNSINNSNMNSN 691
               ||||  :                 ||  :|  :|        ||::||
Db        713 ENIINANANGGAGPGAGVAVGSSIPNVRNIKSSRGLPSAASNLDSN 758
```

RESULT 22

```
KF3B_HUMAN
ID   KF3B_HUMAN     STANDARD;      PRT;   747 AA.
AC   O15066;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin
DE   motor 3B) (HH0048).
GN   KIF3B OR KIAA0359.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21638749; PubMed=11780052;
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
RN   [3]
RP   IDENTIFICATION IN A COMPLEX WITH SMC3 AND KIFAP3B.
RX   MEDLINE=98175913; PubMed=9506951;
RA   Shimizu K., Shirataki H., Honda T., Minami S., Takai Y.;
RT   "Complex formation of SMAP/KAP3, a KIF3A/B ATPase motor-associated
RT   protein, with a human chromosome-associated polypeptide.";
RL   J. Biol. Chem. 273:6591-6594(1998).
CC   -!- FUNCTION: Involved in tethering the chromosomes to the spindle
CC       pole and in chromosome movement. Microtubule-based anterograde
CC       translocator for membranous organelles. Plus end-directed
CC       microtubule sliding activity in vitro (By similarity).
CC   -!- SUBUNIT: Heterodimer of KIF3A and KIF3B (By similarity). Interacts
CC       with the SMC3 subunit of the cohesin complex.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       II SUBFAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB002357; BAA20815.1; -.
DR   EMBL; AL121897; CAC16425.1; -.
DR   HSSP; P17119; 3KAR.
DR   Genew; HGNC:6320; KIF3B.
DR   MIM; 603754; -.
DR   GO; GO:0005873; C:plus-end kinesin complex; TAS.
DR   GO; GO:0003777; F:microtubule motor activity; TAS.
DR   GO; GO:0008574; F:plus-end-directed kinesin ATPase activity; TAS.
DR   GO; GO:0008089; P:anterograde axon cargo transport; TAS.
DR   GO; GO:0007368; P:determination of left/right asymmetry; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Neurone.
FT   DOMAIN        1    345       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      346    579       COILED COIL (BY SIMILARITY).
FT   DOMAIN      580    747       GLOBULAR.
FT   NP_BIND      96    103       ATP (POTENTIAL).
FT   DOMAIN      386    393       POLY-GLY.
FT   DOMAIN      394    406       POLY-GLU.
FT   DOMAIN      723    730       POLY-SER.
SQ   SEQUENCE   747 AA;  85125 MW;  97FA4573AFA87023 CRC64;

  Query Match             6.9%;  Score 472.5;  DB 1;  Length 747;
  Best Local Similarity   32.0%;  Pred. No. 5.1e-14;
  Matches  138;  Conservative   74;  Mismatches  148;  Indels   71;  Gaps   14;

Qy          2 NSKIKVVVRKRPLSELEK-----KKKDSDI----ITVKNNCTLYIDEPRYKVDMTKYIER 52
                : ::|||| ||::  ||      |  | |:     ::|||          : |:
Db          7 SESVRVVVRCRPMNGKEKAASYDKVVDVDVKLGQVSVKNPKGTAHEMPK---------- 55

Qy         53 HEFIVDKVFDDTVDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGSQP 110
                |  | |:|      | :|: | :||:  : :   || :    |||||||:||||| | :
Db         56 -TFTFDAVYDWNAKQFELYDETFRPLVDSVLQGFNGTI---FAYGQTGTGKTYTMEGIR- 110

Qy        111 YGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQK--RKMVAALE 168
                | :  |:   :    ||| ::    ::     :  |: |||   :: ||| |   | :  |
Db        111 -GDPEKRGVIPNSFDHIFTHIS-RSQNQQYLVRASYLEIYQEEIRDLLSKDQTKRLELKE 168

Qy        169 NGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK---- 224
                 | ||||          : :|:   |     |     | :|  : |: ||||||  | ::
Db        169 RPDTGVYVKDLSSFVTKSVKEIEHVMNVGNQNRSVGATNMNEHSSRSHAIFVITIECSEV 228

Qy        225 --DINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM-DSDK 281
                  |   :  :||:  :||||||| | | :| :: : :     || || ||     | |: |
Db        229 GLDGENHIRVGKLNLVDLAGSERQAKTGAQGERLK-EATKINLSLSALGNVISALVDGKS 287

Qy        282 NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCI 341
                |||:|||:||::|:|    | :|::|:||: |         |:|| ||||::| || |||    :
Db        288 THIPYRDSKLTRLLQDSLGGNAKTVMVANVGPASYNVEETLTTLRYANRAKNIKNKPR-V 346

Qy        342 NEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNIL 401
                ||            | | : :       |      :|:                   ::|: :|
Db        347 NE-----------DPKDALLREFQEEIARLKA------------------QLEKRSIG 375

Qy        402 KNKSFDKPREG 412
                : |   :| |||
Db        376 RRKRREKRREG 386


RESULT 23
KF3B_MOUSE
ID   KF3B_MOUSE     STANDARD;      PRT;   747 AA.
AC   Q61771;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin
DE   motor 3B).
GN   KIF3B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=Brain;
RX   MEDLINE=96032268; PubMed=7559760;
RA   Yamazaki H., Nakata T., Okada Y., Hirokawa N.;
RT   "KIF3A/B: a heterodimeric kinesin superfamily protein that works as a
RT   microtubule plus end-directed motor for membrane organelle
RT   transport.";
RL   J. Cell Biol. 130:1387-1399(1995).
CC   -!- FUNCTION: Involved in tethering the chromosomes to the spindle
CC       pole and in chromosome movement. Microtubule-based anterograde
CC       translocator for membranous organelles. Plus end-directed
CC       microtubule sliding activity in vitro (By similarity).
CC   -!- SUBUNIT: Interacts with the SMC3 subunit of the cohesin
CC       complex (By similarity). Heterodimer of KIF3A and KIF3B.
```

```
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       II SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D26077; BAA05070.1; -.
DR   PIR; A57107; A57107.
DR   HSSP; P17119; 3KAR.
DR   MGD; MGI:107688; Kif3b.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Neurone.
FT   DOMAIN        1    345       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      346    579       COILED COIL (BY SIMILARITY).
FT   DOMAIN      580    747       GLOBULAR.
FT   NP_BIND      96    103       ATP (POTENTIAL).
FT   DOMAIN      386    393       POLY-GLY.
FT   DOMAIN      394    405       POLY-GLU.
FT   DOMAIN      723    730       POLY-SER.
SQ   SEQUENCE   747 AA;  85288 MW;  FA369A4190EC8B47 CRC64;

  Query Match             6.9%;  Score 472.5;  DB 1;  Length 747;
  Best Local Similarity   32.0%;  Pred. No. 5.1e-14;
  Matches  138;  Conservative   74;  Mismatches  148;  Indels   71;  Gaps   14;

Qy          2 NSKIKVVVRKRPLSELEK-----KKKDSDI----ITVKNNCTLYIDEPRYKVDMTKYIER 52
                :  ::||||  ||::   ||        |    |  |:      ::|||        :  |:
Db          7 SESVRVVVRCRPMNGKEKAASYDKVVDVDVKLGQVSVKNPKGTSHEMPK---------- 55

Qy         53 HEFIVDKVFDDTVDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGSQP 110
                 |  |  |:|     |:|:  |  :||:  :  :   ||  :     |||||||:||||||  |  :
Db         56 -TFTFDAVYDWNAKQFELYDETFRPLVDSVLQGFNGTI---FAYGQTGTGKTYTMEGVR- 110

Qy        111 YGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQK--RKMVAALE 168
                |  :   |:     :     |||  ::       ::     |: |||   :: |||  |    :    |
Db        111 -GDPEKRGVIPNSFDHIFTHIS-RSQNQQYLVRASYLEIYQEEIRDLLSKDQTKRLELKE 168

Qy        169 NGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK---- 224
                      |  ||||            :  :|:     |    |     | :|    :  |:  ||||||     |  ::
Db        169 RPDTGVYVKDLSSFVTKSVKEIEHVMNVGNQNRSVGATNMNEHSSRSHAIFVITIECSEV 228

Qy        225 --DINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM-DSDK 281
                   |     :   :||:   :||||||   |   |  :|   ::  :  :      ||   ||  ||     |  :  |
Db        229 GLDGENHIRVGKLNLVDLAGSERQAKTGAQGERLK-EATKINLSLSALGNVISALVDGKS 287

Qy        282 NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCI 341
                 |||:|||:||::|:|     |   :|::|:||:|   |        |:||   ||||:: |  ||   |||          :
Db        288 THIPYRDSKLTRLLQDSLGGNAKTVMVANVGPASYNVEETLTTLRYANRAKNIKNKPR-V 346

Qy        342 NEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNIL 401
                ||                      |   |  :   :           |    :|:                              ::|:  :|
Db        347 NE------------DPKDALLREFQEEIARLKA-------------------QLEKRSIG 375

Qy        402 KNKSFDKPREG 412
                :  |    :|  |||
Db        376 RRKRREKRREG 386


RESULT 24
KLP3_SCHPO
ID   KLP3_SCHPO     STANDARD;      PRT;   554 AA.
AC   Q9US60; Q9US61;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein 3 (Kinesin-related protein 1).
GN   KLP3 OR KRP1 OR SPAC1834.07.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RC   STRAIN=972;
RX   MEDLINE=20109189; PubMed=10641037;
RA   Brazer S.-C.W., Williams H.P., Chappell T.G., Cande W.Z.;
RT   "A fission yeast kinesin affects Golgi membrane recycling.";
RL   Yeast 16:149-166(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=22127603; PubMed=12132578;
RA   Jeong J.W., Rhee D.K., Cho S.Y., Hae K.L., Kim D.U., Won M., Kim H.B.;
RT   "Cloning and characterization of the kinesin-related protein, Krp1p,
RT   in Schizosaccharomyces pombe.";
RL   Mol. Cells 13:389-398(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: CYTOPLASMIC MOTOR THAT COULD PLAY A ROLE IN GOLGI
CC       MEMBRANE RECYCLING.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF154055; AAF14525.1; -.
DR   EMBL; AF156966; AAF22609.1; -.
DR   EMBL; AF247188; AAF81205.1; -.
DR   EMBL; AL157734; CAB75775.1; -.
DR   PIR; T50118; T50118.
DR   HSSP; P33176; 1BG2.
```

```
DR   GeneDB_SPombe; SPAC1834.07; -.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   ATP-binding; Coiled coil; Motor protein.
FT   DOMAIN        1    332       KINESIN-MOTOR (BY
FT                                SIMILARITY).
FT   DOMAIN      446    473       COILED COIL (POTENTIAL).
FT   NP_BIND      84     91       ATP (POTENTIAL).
FT   NP_BIND     233    240       ATP (POTENTIAL).
FT   CONFLICT     57     57       Q -> L (IN REF. 1).
SQ   SEQUENCE   554 AA;  61938 MW;  B4533BD095C49CBD CRC64;

  Query Match             6.8%;  Score 465.5;  DB 1;  Length 554;
  Best Local Similarity   25.8%;  Pred. No. 7.4e-14;
  Matches  157;  Conservative  121;  Mismatches  191;  Indels  139;  Gaps   22;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFD 62
                :|||| | || ::||:    :::|    |: |::|:   |          |: |:||
Db          2 TSIKVVCRIRPTNQLEQDLGGNNVIYPLNDSTVHIETSDYS---------GNFVFDRVFH 52

Qy         63 DTVDNFTVYENTIKPLIIDLY--ENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD--TPG 118
                :    ::  :|:  : ||:     || |     ||||||||||||:|  :    :  :  ||
Db         53 PSSTQNDIFSYSIESTVDDLFLGYNGTV---LAYGQTGSGKTYTMMGIENNFEKEGMTPR 109

Qy        119 IFQYAAGDIFTFLNIYDKDNTK--GIFISFYEIYCGKLYDLL-QKRKMVAALENGKKEVV 175
                : :          |  |  |  :|     : :|: |||  |::|||  :|    :    |:  :  |
Db        110 MLRR------IFDKIRDSPSTTEYEVKVSYMEIYMEKIHDLLSEKNDRLTVHEDKLQGVY 163

Qy        176 VKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINK---NTSL 232
                |: || :  | ::  |  :   :   |:   |  :    | |  :|||||:|  :::    :      |
Db        164 VQGLKTIYVSSETEALDILNKGMGSRAVASTSMNAQSSRSHSIFVLEVVQTDTESGETRR 223

Qy        233 GKIAFIDLAGSERGADTVSQN---KQTQTDGANINRSLLALKECIRAM-DSDKNHIPFRD 288
                |::  :||||||     :| ::     ||   :     ||||  |    |  ::  ||   :|:|:||
Db        224 GRLFLVDLAGSE----SVGKSGAVGQTLEEAKKINRSLSTLGMVINSLTDSKLSHVPYRD 279

Qy        289 SELTKVLRDIFVGKSKSIMIANISPTIISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
                |:||::|::    ||::  :| || ||      :||:|||:   |  |:  |||:    :| |
Db        280 SKLTRILKESLGGNSRTTLIINCSPDSYNATETLSTLRFGHRAKSIKNKAV-VNSE---- 334

Qy        349 TERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDK 408
                  :|: :  |                      :   ::  :   :   :   :||:  ::  ||   |
Db        335 ---LSVDEMK--------------RQLYIYKDALSRCVCGARINNNLDYNNCHSN----- 372

Qy        409 PREGFTSTFGKYS-SLNDIDKIKKNKKKKGLINYKSTLYNDNTINKKHNNNNNNNDNNDN 467
                         |::| :|::: :  |  |  |:  |       :   :  ||  :        :|||  :::
Db        373 ------VWSGEHSLTLSNLAE-KSNLKEAEI-----IQGNRTIQE-------SNNDRDES 413

Qy        468 NNDNNNNNNNDSSSM--------------------------------------VNNMIN 488
                  :  :  :|  ||  |:                                              |   :   |
Db        414 TVASIHRHNFDSDSINRLYAEAQLELKQRDGVLSSTKQQLSDLMTALGDAQERYVELVKN 473

Qy        489 HMINNNINNNINVNN-------NNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCH 541
                 | :|:|:  |  ::|:                :  |  :  ||   ||  |  :          |:||
Db        474 HRVNSNLTANNSLNDKPGFTIEQKDKNFSINNERNNFLQK----------LSTLDSSLAA 523

Qy        542 LNNNDKSI 549
                | |   :  :
Db        524 LVNVQRKL 531


RESULT 25
P3K3_DICDI
ID   P3K3_DICDI     STANDARD;      PRT;  1585 AA.
AC   P54675;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphatidylinositol 3-kinase 3 (EC 2.7.1.137) (PI3-kinase)
DE   (PtdIns-3-kinase) (PI3K) (Fragment).
GN   PIKC OR PIK3.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE=96009592; PubMed=7565716;
RA   Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT   "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
RT   discoideum: biological roles of putative mammalian p110 and yeast
RT   Vps34p PI 3-kinase homologs during growth and development.";
RL   Mol. Cell. Biol. 15:5645-5656(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol = ADP +
CC       1-phosphatidyl-1D-myo-inositol 3-phosphate.
CC   -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U23478; AAA85723.1; -.
DR   PIR; T18274; T18274.
DR   DictyDb; DD01101; pikC.
DR   InterPro; IPR000403; PI3_PI4_kinase.
DR   InterPro; IPR002420; PI3K_C2.
DR   InterPro; IPR000341; PI3K_ras_bind.
DR   InterPro; IPR001263; PI3Ka.
DR   Pfam; PF00454; PI3_PI4_kinase; 1.
DR   Pfam; PF00792; PI3K_C2; 1.
DR   Pfam; PF00794; PI3K_rbd; 1.
DR   Pfam; PF00613; PI3Ka; 1.
DR   SMART; SM00142; PI3K_C2; 1.
DR   SMART; SM00144; PI3K_rbd; 1.
DR   SMART; SM00145; PI3Ka; 1.
DR   SMART; SM00146; PI3Kc; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR   PROSITE; PS50290; PI3_4_KINASE_3; 1.
KW   Transferase; Kinase; Multigene family; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       58     84       POLY-ASN.
FT   DOMAIN       98    110       POLY-ASN.
FT   DOMAIN      200    226       POLY-ASN.
FT   DOMAIN      239    254       POLY-ASN.
FT   DOMAIN      345    378       POLY-ASN.
FT   DOMAIN      383    390       POLY-ASN.
FT   DOMAIN      484    488       POLY-SER.
FT   DOMAIN      720    737       POLY-GLN.
FT   DOMAIN     1569   1575       POLY-ASN.
FT   DOMAIN     1221   1484       PI3K/PI4K.
FT   DOMAIN     1510   1539       5 X 5 AA APPROXIMATE REPEATS.
FT   REPEAT     1510   1514       1.
FT   REPEAT     1515   1519       2.
FT   REPEAT     1520   1524       3.
FT   REPEAT     1530   1534       4.
FT   REPEAT     1535   1539       5.
FT   DOMAIN     1547   1560       7 X 2 AA TANDEM REPEATS OF K-E.
SQ   SEQUENCE   1585 AA;  180421 MW;  4689B620D2484961 CRC64;

  Query Match             6.8%;  Score 465.5;  DB 1;  Length 1585;
  Best Local Similarity   24.5%;  Pred. No. 2.3e-13;
  Matches  181;  Conservative   93;  Mismatches  212;  Indels  253;  Gaps   29;

Qy        404 KSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNND 463
                :::||      |   |  ::|:|       ||  |:   |   |       |    |  | :||||||||:
Db         15 ENYDKLES--TEQFKQWSNL-----IKNIKENSLNNSNIYLTIPTTQNLINNNNNNNNNN 67
```

```
Qy        464 NNDNNNDNNNNNNNDSSSMVNNMINHMINNNI--------NNNINVNNNNNNNNNNNNSH 515
              ||:|||:||||||              |||:        |  |||:||||| | |
Db         68 NNNNNNNNNNNNNN-------------NNNVIIPSASTENKEENDNNNSNNNNNINLSP 113

Qy        516 NNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKK 575
              :: :      | |  : ::   :  :     |:     |   |:  |  |:: |
Db        114 DSSI---------TKDINITENKITEIKTTET-----KETSTGTSPLEKSPSKGFIISPK 159

Qy        576 K------------NNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILY 623
              |             ||: :  :: |  :              |   |  :  :: |
Db        160 KPEEENEIEGETINNIAITNYTQGPSML------------TLMKKKLENIKKN------- 200

Qy        624 ESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGV-------INIINNSNVNSIN 676
                | :| |||   | |::::      | |  | |||:       ::  |:| |: |
Db        201 --NNNNNNNGNG----NNNSNNN------NSNSNNNNNGISPSSSPPSHLNGNNNNNNSN 248

Qy        677 NSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNM 736
              |:| |:  |:  ||      :  |||      ::         :|   : :| | | |  ||
Db        249 NTNSNNTTNATTNSVGFSITMTNSNSLSVSKRMNKFKSWTSSKPTSSSIGF-ASSPQNNG 307

Qy        737 YP------------------------------NITNNNNNNNNNNNNNNNIDVENYN 764
              |                              :| |  |||||||||||||     | |
Db        308 KPLNISGSSRFFTSRQDSKIDLLKSPSSSPPTQSDIFNENNNNNNNNNNNNN----NNN 363

Qy        765 NRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKN 824
              | :  ||                   :|| | :| : || | ||                ||:|
Db        364 NNNNNNN---------------NNNNNNNNEELINNNNNNN---------------NDEN 393

Qy        825 YLID-----LNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNS 879
              | |:     |     :|:|  :      |:: ||: :  || |   :
Db        394 YKIEETEESLKELLEKEKLENEEREKILKERNEIDNLKKKNHLSK-------------- 438

Qy        880 PRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSN------------NNNSSSS 926
                      |   | |::     | :|   ::| | ::|               :||||
Db        439 --------GYFMRACNASNDD--GLEEEDIPLQDEHWETNVIVLLPCRHHVKVPGSSSSS 488

Qy        927 SSSNNNI---------YNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYT 977
                      :    |:  |: |          ||:|                :  |  |:::
Db        489 IDSIRQLAWASGKMQGHLNLEKDEKF----------FTLR-----------WCNKDVVFD 527

Qy        978 INSLNDYMSNTLLHFKEKYTYPTLS-TN-------EDIYNKEMEGKHIRLDDQDKYDDND 1029
                    |      |: :  |  ||   ||         ||  ||      || |      :
Db        528 ----QDTPLGHLIQYNLNYNNPTQKPTNIKLELVLEDELCKE------RLVDLQSLE--I 575

Qy       1030 NNNVDNNNKNNVDNNVDNN 1048
              ||    :   |:::|: :  |
Db        576 NNGRPSIWKSHIDDVLSFN 594
```

RESULT 26

```
CIN8_YEAST
ID   CIN8_YEAST     STANDARD;      PRT;  1038 AA.
AC   P27895;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin-like protein CIN8.
GN   CIN8 OR KSL2 OR YEL061C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=92317149; PubMed=1618897;
RA   Hoyt M.A., He L., Loo K.K., Saunders W.S.;
RT   "Two Saccharomyces cerevisiae kinesin-related gene products required
RT   for mitotic spindle assembly.";
RL   J. Cell Biol. 118:109-120(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169868;
RA   Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA   Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA   Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA   Hunicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA   Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA   Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA   Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL   Nature 387:78-81(1997).
RN   [3]
RP   SEQUENCE OF 1021-1038 FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=95172238; PubMed=7867803;
RA   Rousselet G., Simon M., Ripoche P., Buhler J.M.;
RT   "A second nitrogen permease regulator in Saccharomyces cerevisiae.";
RL   FEBS Lett. 359:215-219(1995).
RN   [4]
RP   CHARACTERIZATION.
RC   STRAIN=S288c;
RX   MEDLINE=92354062; PubMed=1643659;
RA   Saunders W.S., Hoyt M.A.;
RT   "Kinesin-related proteins required for structural integrity of the
RT   mitotic spindle.";
RL   Cell 70:451-458(1992).
CC   -!- FUNCTION: REQUIRED FOR ASSEMBLY OF THE MITOTIC SPINDLE. INTERACT
CC       WITH SPINDLE MICROTUBULES TO PRODUCE AN OUTWARDLY DIRECTED
CC       FORCE ACTING UPON THE POLES. FOLLOWING SPINDLE ASSEMBLY, CIN8 AND
CC       KIP1 APPARENTLY ACT TO OPPOSE A FORCE THAT DRAWS SEPARATED POLES
CC       BACK TOGETHER. THIS FORCE SEEMS TO BE MEDIATE BY KAR3.
CC   -!- SUBCELLULAR LOCATION: SPINDLE MICROTUBULES THAT LIE BETWEEN THE
CC       POLES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z11859; CAA77885.1; -.
DR   EMBL; M90522; AAA34496.1; -.
DR   EMBL; U18795; AAB65026.1; -.
DR   EMBL; X79105; CAA55722.1; -.
DR   PIR; S50528; B42641.
DR   HSSP; P17119; 3KAR.
DR   SGD; S0000787; CIN8.
DR   GO; GO:0000092; P:mitotic anaphase B; IMP.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW   Mitosis; Cell cycle.
FT   DOMAIN       72    553       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      554    657       COILED COIL (POTENTIAL).
FT   DOMAIN      904    942       COILED COIL (POTENTIAL).
FT   NP_BIND     166    173       ATP (BY SIMILARITY).
FT   CONFLICT    254    254       D -> A (IN REF. 1).
FT   CONFLICT    831    831       Q -> H (IN REF. 1).
SQ   SEQUENCE   1038 AA;  117999 MW;  3A1FD7003EF89FBC CRC64;

  Query Match             6.8%;  Score 463;  DB 1;  Length 1038;
  Best Local Similarity   20.4%;  Pred. No. 1.9e-13;
  Matches  247;  Conservative  172;  Mismatches  374;  Indels  416;  Gaps   47;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKN---------NCTLYIDEPRYKVDMTKYIERHEF 55
              | | || |  :| |   | | :: | :          | |              :|  :   :
```

```
Db         75 ITVAVRCRGRNEREISMKSSVVVNVPDITGSKEISINTT-------GDTGITAQMNAKRY 127

Qy         56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPY--GQ 113
              |||||       :::       ||  |  :  |  |:    ||  | :||||||  | :    |:
Db        128 TVDKVFGPGASQDLIFDEVAGPLFQD-FIKGYNCTVLVYGMTSTGKTYTMTGDEKLYNGE 186

Qy        114 -SDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYD--------------- 156
               ||  ||       :|  |  :    |     :  || |:|  :| |
Db        187 LSDAAGIIPRVLLKLFDTLELQQNDYV--VKCSFIELYNEELKDLLDSNSNGSSNTGFDG 244

Qy        157 -------------------------------------------LLQKRKMVAALENGK 171
                                                         ||:||    :| |
Db        245 QFMKKLRIFDSSTANNTTSNSASSSRSNSRNSSPRSLNDLTPKAALLRKRLRTKSLPNTI 304

Qy        172 KE-------------------------------------------VVVKDLKILRV 184
              |:                                           : :::|:   :
Db        305 KQQYQQQQAVNSRNNSSSNSGSTTNNASSNTNTNTNNGQRSSMAPNDQTNGIYIQNLQEFHI 364

Qy        185 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDL--KDINKNTSLGKIAFIDLAG 242
               | :  :  |:  |::          || |||||.| ||  |  |    ::    :  |:  :||||
Db        365 TNAMEGLNLLQKGLKHRQVASTKMNDFSSRSHTIFTITLYKKHQDELFRISKMNLVDLAG 424

Qy        243 SERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGK 302
              ||    :  :  |::  :    | :||:|||  |     |  |:      |||||:|:||::|:|   |
Db        425 SENINRSGALNQRAKEAG-SINQSLLTLGRVINALVDKSGHIPFRESKLTRLLQDSLGGN 483

Qy        303 SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEM 362
              :|: :|| |||       |:|  :|| |:|:  || |||                     ::
Db        484 TKTALIATISPAKVTSEETCSTLEYASKAKNIKNK---------------------PQL 521

Qy        363 NASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSS 422
               :   :::::|          |    :  ||    ::::|  ||    :||  :   |
Db        522 GSFIMKDILVK---------NITMELAKI-----KSDLLSTKS----KEGIYMSQDHYKN 563

Qy        423 LN-DIDKIK------KNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNN 475
              || |:: |      |  : : |  :  |   :  |    :|:   :   |   ::
Db        564 LNSDLESYKNEVQECKREIESLTSKNALLVKDKLKSKETIQSQNCQIES---------- 612

Qy        476 NNDSSSMVNNMINHM---INNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDF 532
                        :   |:|:  ::    | :::  ||                     |:
Db        613 -------LKTTIDHLRAQLDKQHKTEIEISDFNNKLQK--------------LTEVMQM 650

Qy        533 SSLDDMNCHLNNNDK---SIFLHKKNLRDNIKLK-NRSSCDNIMNKK--KNNLHLARHSV 586
               :  |      |:  |  |    |    | |:  :  |:  |    ::|:  :    :  ||  : :: |
Db        651 ALHDYKKRELDLNQKFEMHITKEIKKLKSTLFLQLNTMQQESILQETNIQPNLDMIKNEV 710

Qy        587 GSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHD 646
                ||:    :    :|    :  :  |:  : :||
Db        711 ---LTLMR--TMQEKAELMYKDCVKKILNESPK------------------------- 738

Qy        647 ISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISD 706
                              |::                :   |:   ::      ||     |    ::::||
Db        739 ---------------FFNVV--------------IEKIDIIRVDFQKFYK---NIAENLSD 767

Qy        707 VQIRYVNEMDTSNKNND------NIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDV 760
              :            | :||:        | ||          |  :  :-  |  : ::    |      |  |
Db        768 I----------SEENNNMKQYLKNHFF------KNNHQELLNRHVDSTYENIEKRTNEFV 811

Qy        761 ENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNL 820
              ||:           |  |        |  ||      :||:  |    :      ::   |:
Db        812 ENFKK-----------VLNDH----LDENKKLI-MQNLTTATS---AVIDQEMDL----- 847

Qy        821 NDKNYLIDLNNKEQKDKN----IHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGN 876
                              |  | :|      |: ||:         |:|                :  |  ::|  |
Db        848 --------FEPKRVKWENSFDLINDCDS----MNNEFYNSMAATLSQIKSTVDTSSNSMN 895

Qy        877 NNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSN----NN 932
               :    ||  :   |   :|  :|||     |:: :        |  |:  |   ||  :|: |:     ::
Db        896 ESISVMKGQVEESENAISLLKNNTKFNDQFEQLINKHNMLKDNIKNSITSTHSHITNVDD 955

Qy        933 IYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHF 992
              ||||                           ||   |  |  |  :            |:   |  |    :|
Db        956 IYNTIE------------------------NIMKN-YGNKE-----NATKDEMIENIL-- 983

Qy        993 KEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNV---------DN 1043
              ||    |  ||         |:|      :||                 :|:::|:  :|               |
Db        984 KE---IPNLS-------KKMP---LRL-----------SNINSNSVQSVISPKKHAIEDE 1019

Qy       1044 NVDNNNVDN 1052
              |   :  ||||
Db       1020 NKSSENVDN 1028
```

```
RESULT 27
KF3A_HUMAN
ID   KF3A_HUMAN     STANDARD;      PRT;   702 AA.
AC   Q9Y496;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein KIF3A (Microtubule plus end-directed kinesin
DE   motor 3A).
GN   KIF3A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=20018159; PubMed=10548469;
RA   Whitehead J.L., Wang S.Y., Bost-Usinger L., Hoang E., Frazer K.A.,
RA   Burnside B.;
RT   "Photoreceptor localization of the KIF3A and KIF3B subunits of the
RT   heterotrimeric microtubule motor kinesin II in vertebrate retina.";
RL   Exp. Eye Res. 69:491-503(1999).
CC   -!- FUNCTION: MICROTUBULE-BASED ANTEROGRADE TRANSLOCATOR FOR
CC       MEMBRANOUS ORGANELLES. PLUS END-DIRECTED MICROTUBULE SLIDING
CC       ACTIVITY IN VITRO.
CC   -!- SUBUNIT: HETERODIMER OF KIF3A AND KIF3B.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       II SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF041853; AAC72294.1; -.
DR   HSSP; P17119; 3KAR.
DR   Genew; HGNC:6319; KIF3A.
DR   MIM; 604683; -.
DR   GO; GO:0006996; P:organelle organization and biogenesis; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Neurone.
FT   DOMAIN        1    350       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      351    587       COILED COIL (BY SIMILARITY).
FT   DOMAIN      588    702       GLOBULAR.
FT   NP_BIND     100    107       ATP (BY SIMILARITY).
FT   DOMAIN      443    446       POLY-GLU.
SQ   SEQUENCE   702 AA;  80337 MW;  24507D1EC6D540DB CRC64;

  Query Match             6.7%;  Score 460;  DB 1;  Length 702;
  Best Local Similarity   29.8%;  Pred. No. 1.7e-13;
  Matches  147;  Conservative   84;  Mismatches  192;  Indels   70;  Gaps   15;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHE----FIVDKV 60
              :||||| |||:| ||          ::|            ||| |  : : |     :|    |  | |
```

```
Db         15 VKVVVRCRPLNEREKSMCYKQAVSV--------DEMRGTITVHKTDSSNEPPKTFTFDTV 66

Qy         61 FDDTVDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPG 118
              |         || |:|:|  : |  || :  ||||||:|||:|| | :      :  |
Db         67 FGPESKQLDVYNLTARPIIDSVLEGYNGTI---FAYGQTGTGKTFTMEGVRAI--PELRG 121

Qy        119 IFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQK--RKMVAALENGKKEVVV 176
              |  :   ||  :    :  |      : :|: |||  :: |||  |    : :|  |     | :
Db        122 IIPNSFAHIFGHIAKAEGDTRFLVRVSYLEIYNEEVRDLLGKDQTQRLAVKERPDVGVYI 181

Qy        177 KDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK------DINKNT 230
              |||     |   :::   |   |      | :|   | :  |: ||||||   | ::      | |  :
Db        182 KDLSAYVVNNADDMDRIMTLGHKNRSVGATNMNEHSSRSHAIFTITIECSEKGIDGNMHV 241

Qy        231 SLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM-DSDKNHIPFRDS 289
               :||: :||||||  | |  :  :: :  :   || ||   |    | |: |     |:|:|:|
Db        242 RMGKLHLVDLAGSERQAKTGATGQRLK-EATKINLSLSTLGNVISALVDGKSTHVPYRNS 300

Qy        290 ELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEED---- 345
              :||::|:|    | ||::| |||  |        ::|::||||::| || |||:  |||:
Db        301 KLTRLLQDSLGGNSKTMMCANIGPADYNYDETISTLRYANRAKNIKNKAR-INEDPKDAL 359

Qy        346 ----DTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKIND-----KIE 396
                    | :     :| |:: | |           | :  :    |:: :     |
Db        360 LRQFQKEIEELKKKLEEGEEISGSDI-----------SGSEEDDEEGEVGEDGEKRKKR 408

Qy        397 RNNILKNK-----------SFDKPREGFTSTFG-----KYSSLNDIDKIKKNKKKGLINY 440
              |  | | |               |: |:          :  :  :::| :|:  |   :
Db        409 RIQIGKKKVSPDKMIEMQAKIDEERKALETKLDMEEEERNKARAELEKREKDLLKAQQEH 468

Qy        441 KSTLYNDNTINKK 453
              :| |   : : ||
Db        469 QSLLEKLSALEKK 481


RESULT 28
KL61_DROME
ID   KL61_DROME     STANDARD;      PRT;  1066 AA.
AC   P46863; Q8T0A6; Q9W0I8;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Bipolar kinesin KRP-130 (Kinesin-like protein Klp61F).
GN   KLP61F OR KLP2 OR CG9191.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=94043448; PubMed=8227131;
RA   Heck M.M.S., Pereira A., Pesavento P.A., Yannoni Y., Spradling A.C.,
RA   Goldstein L.S.B.;
RT   "The kinesin-like protein KLP61F is essential for mitosis in
RT   Drosophila.";
RL   J. Cell Biol. 123:665-679(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   REVISIONS.
RC   STRAIN=Berkeley;
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN   [5]
RP   SEQUENCE OF 228-357 FROM N.A.
RC   STRAIN=DP CN BW;
RX   MEDLINE=92020874; PubMed=1924306;
RA   Stewart R.J., Pesavento P.A., Woerpel D.N., Goldstein L.S.B.;
RT   "Identification and partial characterization of six members of the
RT   kinesin superfamily in Drosophila.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:8470-8474(1991).
RN   [6]
RP   PARTIAL SEQUENCE, AND IDENTIFICATION AS KRP-130.
RX   MEDLINE=97078747; PubMed=8918872;
RA   Kashina A.S., Scholey J.M., Leszyk J.D., Saxton W.M.;
RT   "An essential bipolar mitotic motor.";
RL   Nature 384:225-225(1996).
CC   -!- FUNCTION: IMPORTANT ROLE IN MITOTIC DIVIDING CELLS. MICROTUBULE
CC       MOTOR REQUIRED FOR SPINDLE BODY SEPARATION. SLOW PLUS-END DIRECTED
CC       MICROTUBULE MOTOR CAPABLE OF CROSS-LINKING AND SLIDING APART
CC       ANTIPARALLEL MICROTUBULES, THEREBY PUSHING APART THE ASSOCIATED
CC       SPINDLE POLES DURING SPINDLE ASSEMBLY AND FUNCTION.
CC   -!- SUBUNIT: Homotetramer.
CC   -!- DEVELOPMENTAL STAGE: SPECIFICALLY EXPRESSED IN PROLIFERATING
CC       TISSUES DURING EMBRYONIC AND LARVAL DEVELOPMENT.
CC   -!- PTM: PHOSPHORYLATION DURING MITOSIS AT THR-933 CONTROLS THE
```

```
CC       ASSOCIATION OF KLP61F WITH THE SPINDLE APPARATUS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U01842; AAA03718.1; -.
DR   EMBL; AE003471; AAF47458.2; -.
DR   EMBL; AY069442; AAL39587.1; -.
DR   EMBL; M74428; AAA28655.1; -.
DR   PIR; A48669; A48669.
DR   HSSP; P17119; 3KAR.
DR   FlyBase; FBgn0004378; Klp61F.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005871; C:kinesin complex; IDA.
DR   GO; GO:0003774; F:motor activity; IDA.
DR   GO; GO:0007100; P:centrosome separation; IGI.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW   Mitosis; Cell cycle; Phosphorylation.
FT   DOMAIN       17    361       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      362    462       COILED COIL (POTENTIAL).
FT   DOMAIN      540    569       COILED COIL (POTENTIAL).
FT   DOMAIN      639    738       COILED COIL (POTENTIAL).
FT   DOMAIN      808    875       COILED COIL (POTENTIAL).
FT   DOMAIN      889    918       COILED COIL (POTENTIAL).
FT   NP_BIND     103    110       ATP (POTENTIAL).
FT   MOD_RES     933    933       PHOSPHORYLATION (BY CDC2) (BY
FT                                SIMILARITY).
FT   VARIANT     595    595       M -> V.
FT   VARIANT     869    869       R -> K.
FT   VARIANT     904    904       H -> Q.
FT   CONFLICT    962    962       L -> Q (IN REF. 1).
FT   CONFLICT    983    983       V -> D (IN REF. 1).
SQ   SEQUENCE   1066 AA;  121163 MW;  363647366EE0721F CRC64;

  Query Match             6.7%;  Score 459;  DB 1;  Length 1066;
  Best Local Similarity   21.0%;  Pred. No. 2.9e-13;
  Matches  180;  Conservative  175;  Mismatches  327;  Indels  176;  Gaps   28;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERH-------- 53
Db         17 NQNIQVYVRVRPLNSRERCIRSAEVVDV--------------VGPREVVTRHTLDSKLTK 62

Qy         54 EFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQ---- 109
Db         63 KFTFDRSFGPESKQCDVYSVVVSPLIEEVL-NGYNCTVFAYGQTGTGKTHTMVGNETAEL 121

Qy        110 --PYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLL---QKRKMV 164
Db        122 KSSWEDDSDIGIIPRALSHLFDELRMMEVEYT--MRISYLELYNEELCDLLSTDDTTKIR 179

Qy        165 AALENGKK-EVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDL 223
Db        180 IFDDSTKKGSVIIQGLEEIPVHSKDDVYKLLEKGKERRKTATTLMNAQSSRSHTVFSI-V 238

Qy        224 KDINKN-------TSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRA 276
Db        239 VHIRENGIEGEDMLKIGKLNLVDLAGSENVSKAGNEKGIRVRETVNINQSLLTLGRVITA 298

Qy        277 MDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKN 336
Db        299 LVDRAPHVPYRESKLTRLLQESLGGRTKTSIIATISPGHKDIEETLSTLEYAHRAKNIQN 358

Qy        337 ---------KSTCINEE----DDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNN- 382
Db        359 KPEVNQKLTKKTVLKEYTEEIDKLKRDLMAARDKNGIYLAEETYGEITLK---LESQNRE 415

Qy        383 -NNKI-----------NRGKINDKIERNNILKNKSFDKPREGFTSTFGKY----SSLNDI 426
Db        416 LNEKMLLLKALKDELQNKEKIFSEVSMSLVEKTQELKKTEENLLNTKGTLLLTKKVLTKT 475

Qy        427 DKIKKNKKKGLINYKST----------------LYNDNT----------------INKKH 454
Db        476 KRRYKEKKELVASHMKTEQVLTTQAQEILAAADLATDDTHQLHGTIERRELDEKIRRSC 535

Qy        455 NNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNN-INNNINVNNNNN------- 506
Db        536 DQFKDRMQDNLEMIGGSLNLYQDQQAALKEQLSQEMVNSSYVSQRLALNSSKSIEMLKEM 595

Qy        507 ---NNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKL- 562
Db        596 CAQSLQDQTNLHNKLIGE----VMKISDQHSQAFVAKLMEQMQQQQLLMSKEIQTNLQVI 651

Qy        563 -----KNRSSCDNIMNK-------------KKNNLHLARHSVGSKLTMFSYDPQKNKDN 603
Db        652 EENNQRHKAMLDSMQEKFATIIDSSLQSVEEHAKQMHKKLEQLGAMSLPDAEELQNLQEE 711

Qy        604 TFFKSNINKMEDNTPKDILYESRNVSNMNG----NVLLGLNKNTHHDISTKDENHNDNKI 659
Db        712 LANERALAQQEDALLESMMMQMEQIKNLRSKNSISMSVHLNK---MEESRLTRNHRIDDI 768

Qy        660 NNGV-----INIINNSNVNS----------------INNSNMNSINNSNMNSNSIYKSNY 698
Db        769 KSGIQDYQKLGIEASQSAQAELTSQMEAGMLCLDQGVANCSMLQVHMKNLNQ----KYEK 824

Qy        699 NSNQSISDVQIRYVNEMD 716
Db        825 ETNENVGSVRVHH-NQVE 841

RESULT 29
KINH_HUMAN
ID   KINH_HUMAN     STANDARD;      PRT;   963 AA.
AC   P33176;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin heavy chain (Ubiquitous kinesin heavy chain) (UKHC).
GN   KIF5B OR KNS1 OR KNS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=92299683; PubMed=1607388;
RA   Navone F., Niclas J., Hom-Booher N., Sparks L., Bernstein H.D.,
RA   McCaffrey G., Vale R.D.;
RT   "Cloning and expression of a human kinesin heavy chain gene:
RT   interaction of the COOH-terminal domain with cytoplasmic microtubules
RT   in transfected CV-1 cells.";
RL   J. Cell Biol. 117:1263-1275(1992).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 1-349.
RX   MEDLINE=96195066; PubMed=8606779;
RA   Kull F.J., Sablin E.P., Lau R., Fletterick R.J., Vale R.D.;
RT   "Crystal structure of the kinesin motor domain reveals a structural
RT   similarity to myosin.";
RL   Nature 380:550-555(1996).
RN   [3]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=94242426; PubMed=7514426;
RA   Niclas J., Navone F., Hom-Booher N., Vale R.D.;
```

```
RT   "Cloning and localization of a conventional kinesin motor expressed
RT   exclusively in neurons.";
RL   Neuron 12:1059-1072(1994).
CC   -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC       PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC   -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC       CHAINS.
CC   -!- SUBCELLULAR LOCATION: UNIFORMLY DISTRIBUTED BETWEEN THE CELL BODY
CC       AND THE PROCESSES IN THE NEURONS.
CC   -!- TISSUE SPECIFICITY: FOUND IN NEWBORN AND ADULT BRAIN, LIVER,
CC       KIDNEY, SPLEEN, HEART, LUNG AND SCIATIC NERVE.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC       TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC       KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC       ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC       DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC       INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC       VESICLES AND MEMBRANOUS ORGANELLES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X65873; CAA46703.1; -.
DR   PIR; A41919; A41919.
DR   PDB; 1BG2; 14-OCT-98.
DR   Genew; HGNC:6324; KIF5B.
DR   MIM; 602809; -.
DR   GO; GO:0005871; C:kinesin complex; TAS.
DR   GO; GO:0004002; F:adenosinetriphosphatase activity; TAS.
DR   GO; GO:0003777; F:microtubule motor activity; TAS.
DR   GO; GO:0007018; P:microtubule-based movement; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; 3D-structure.
FT   DOMAIN        1    256       KINESIN-MOTOR.
FT   DOMAIN      329    914       COILED COIL.
FT   DOMAIN      915    963       GLOBULAR.
FT   NP_BIND      85     92       ATP (BY SIMILARITY).
FT   STRAND        8     15
FT   HELIX        20     25
FT   TURN         26     26
FT   STRAND       29     29
FT   STRAND       32     34
FT   TURN         35     37
FT   STRAND       38     41
FT   TURN         42     43
FT   STRAND       44     47
FT   STRAND       50     52
FT   TURN         54     55
FT   HELIX        58     65
FT   TURN         66     66
FT   HELIX        67     74
FT   TURN         75     76
FT   STRAND       79     84
FT   TURN         87     88
FT   HELIX        91     95
FT   TURN         96     96
FT   STRAND       97     97
FT   TURN         99    100
FT   TURN        102    104
FT   STRAND      105    105
FT   HELIX       107    122
FT   STRAND      126    138
FT   TURN        139    140
FT   STRAND      141    144
FT   TURN        148    149
FT   STRAND      153    153
FT   STRAND      155    157
FT   TURN        159    160
FT   STRAND      163    165
FT   TURN        166    167
FT   STRAND      171    173
FT   HELIX       176    189
FT   TURN        190    193
FT   HELIX       197    203
FT   STRAND      205    216
FT   TURN        217    219
FT   STRAND      222    231
FT   TURN        246    247
FT   HELIX       256    269
FT   TURN        270    271
FT   HELIX       277    279
FT   HELIX       281    285
FT   HELIX       286    288
FT   TURN        289    289
FT   STRAND      295    302
FT   STRAND      305    305
FT   HELIX       306    308
FT   HELIX       309    320
FT   TURN        321    321
SQ   SEQUENCE   963 AA;  109685 MW;  A1FE5760C3250C8B CRC64;

  Query Match             6.6%;  Score 453;  DB 1;  Length 963;
  Best Local Similarity   21.8%;  Pred. No. 4.7e-13;
  Matches  215;  Conservative  171;  Mismatches  421;  Indels  178;  Gaps   34;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 64

Db          9 IKVMCRFRPLNESEVNRGDK-----------YI--AKFQGEDTVVIASKPYAFDRVFQSS 55

Qy         65 VDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGS--QPYGQSDTPGIF 120

Db         56 TSQEQVYNDCAKKIVKDVLEGYNGTI---FAYGQTSSGKTHTMEGKLHDPEGMGIIPRIV 112

Qy        121 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEV-VVKDL 179

Db        113 Q----DIFNYIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLSVHEDKNRVPYVKGC 168

Qy        180 KILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTS---LGKIA 236

Db        169 TERFVCSPDEVMDTIDEGKSNRHVAVTNMNEHSSRSHSIFLINVKQENTQTEQKLSGKLY 228

Qy        237 FIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLR 296

Db        229 LVDLAGSEKVSKTGAEG-AVLDEAKNINKSLSALGNVISALAEGSTYVPYRDSKMTRILQ 287

Qy        297 DIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILD 356

Db        288 DSLGGNCRTTIVICCSPSSYNESETKSTLLFGQRAKTIKN-TVCVNVE------------ 334

Qy        357 SKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNI--LKNKSFDKPREGFT 414

Db        335 --------------------LTAEQWKKKYEKEKEKNKILRNTIQWLENE-LNRWRNGET 373

Qy        415 STFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNND-------- 466

Db        374 VPIDE-----QFDKEKANLEAFTVDKDITLTNDKPATAIGVIGNFTDAERRKCEEEIAKL 428

Qy        467 -NNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYA 525

Db        429 YKQLDDKDEEINQQSQLVEKLKTQMLDQE-----ELLASTRRDQDNMQAELNRLQAENDA 483

Qy        526 FTD--TSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLAR 583

Db        484 SKEEVKEVLQALEEL--AVNYDQKS-----QEVED--KTKEYELLSDELNQKSATL---- 530
```

```
Qy        584 HSVGSKLTMFSYDPQKNKDNT-FFKSNINKMEDNTPKDI--------------------- 621

Db        531 ASIDAEL-------QKLKEMTNHQKKRAAEMMASLLKDLAEIGIAVGNNDVKQPEGTGMI 583

Qy        622 --------LYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVN 673

Db        584 DEEFTVARLYISKMKSEVKTMVKRCKQLESTQTESNKKMEENEKELAACQLRI------- 636

Qy        674 SINNSNMNSINNSNMNSNSIYKSNYNSNQSISD--VQIRYVNEMDTSNKNNDNIFFDAIS 731

Db        637 SQHEAKIKSLTEYLQNVEQKKRQLEESVDALSEELVQLRAQEKVHEMEKEHLNKVQTANE 696

Qy        732 CDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYN-----SHNLFQP 786

Db        697 VKQAVEQQIQSHRETHQKQISSLRDEVEAKAKLITDLQDQNQKMMLEQERLRVEHEKLKA 756

Qy        787 DNNKNTSNIQNINTNKNNQ----------DGNVNYSMNFCHYNLNDKNYLIDLNNKEQKD 836

Db        757 TDQEKSRKLHELTVMQDRREQARQDLKGLEETVAKELQTLH-NLR-KLFVQDLATRVKKS 814

Qy        837 KNIHGCDNNIIQNRNDFEKKKKTNFYNNN--NIVIVNNNMGNNNS------PRMKYGLCG 888

Db        815 AEIDSDDTG-----GSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRL-- 867

Qy        889 SHTSIDNMKNNEMKNNEMKDNEMKD 913

Db        868 -RATAERVKALESALKEAKENASRD 891
```

RESULT 30
KIP3_YEAST

```
ID   KIP3_YEAST     STANDARD;      PRT;   805 AA.
AC   P53086;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein KIP3.
GN   KIP3 OR YGL216W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=97435481; PubMed=9290212;
RA   Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT   "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT   chromosome VII.";
RL   Yeast 13:1077-1090(1997).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       II SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z72739; CAA96933.1; -.
DR   PIR; S64238; S64238.
DR   HSSP; P17119; 3KAR.
DR   SGD; S0003184; KIP3.
DR   GO; GO:0005881; C:cytoplasmic microtubule; IDA.
DR   GO; GO:0005880; C:nuclear microtubule; IDA.
DR   GO; GO:0007019; P:microtubule depolymerization; IMP.
DR   GO; GO:0000071; P:mitotic spindle assembly (sensu Saccharomyces); IGI.
DR   GO; GO:0000132; P:mitotic spindle orientation; IGI.
DR   GO; GO:0000065; P:nuclear migration (sensu Saccharomyces); IMP.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    436       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      449    481       COILED COIL (POTENTIAL).
FT   NP_BIND     192    199       ATP (POTENTIAL).
SQ   SEQUENCE   805 AA;  91090 MW;  6049EF2C13D5C2C7 CRC64;

  Query Match             6.6%;  Score 452.5;  DB 1;  Length 805;
  Best Local Similarity   23.1%;  Pred. No. 4.1e-13;
  Matches  168;  Conservative  112;  Mismatches  224;  Indels  223;  Gaps   23;

Qy         52 RH----EFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLG 107

Db        146 RHGGEIKFVFDKLFDETSSQARVYKETTSPL-LDSVLDGFNSTVFAYGATGCGKTYTVSG 204

Qy        108 SQPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKR---KMV 164

Db        205 T-----PSQPGIIFLAMEELFNKITDLKDEKDFEISLSYLEIYNERIRDLLKPETPSKRL 259

Qy        165 AALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK 224

Db        260 VIREDTQNHIKVANLSYHHPNTVEDVMDLVVQGNINRTTSPTEANEVSSRSHAVLQIHIM 319

Qy        225 DINKNTSL------GKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM- 277

Db        320 QTNKLVDLTSQHTFATLSIIDLAGSERAAATRNRGIRLH-EGANINRSLLALGNCINALC 378

Qy        278 --DSDKN-HIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNF 334

Db        379 LNDGSRSCHIPYRDSKLTRLLKFSLGGNCKTVMIVCISPSSSHYDETLNTLKYANRAKEI 438

Qy        335 KNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDK 394

Db        439 KTK--------------------------------IIRNQQSLSRH------------- 452

Qy        395 IERNNILKNKSFDKPREGFTSTFGKYSSL-----NDIDKIKKNKKKGLINYKSTLYNDNT 449

Db        453 ----------------------VGSYLKMITEQKRQIEELREREEK-MISLKLTKYK--- 486

Qy        450 INKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNN 509

Db        487 LNKE--------------------------KIQLAINECVNR--------------- 502

Qy        510 NNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLK------ 563

Db        503 ----------VQQTYAGVETYQVA----------KTLKSLILCKRRFLQMVKLEVDNLIL 542

Qy        564 --NR------------SSCDNIMNKKKNNLHLAR----------------HSVGSKLTMF 593

Db        543 LFEREESTAAEMQPVISNCRMISGQLYNKIHELEMKFDETDTLSSVIHQVHSIDLNKLRE 602

Qy        594 SYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLG--------------L 639

Db        603 MEDWDETYDLVYLESCLNQISELQRNEILV---NSSIMTEKLMSDPGLNSRFKFLSKWLM 659

Qy        640 NKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMN----SNSIYK 695

Db        660 NRTPNIESIIQDLVHIDEEFESFARTFIANPDSN-FTNTNINIINTTAADLAVPAETLQR 718

Qy        696 SNYNSNQ 702

Db        719 QNFSQKK 725
```

RESULT 31
KLP5_SCHPO

```
ID   KLP5_SCHPO     STANDARD;      PRT;   883 AA.
AC   O14343;
DT   15-JUL-1998 (Rel. 36, Created)
```

```
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein 5.
GN   KLP5 OR SPBC2F12.13.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21963728; PubMed=11967147;
RA   Garcia M.A., Koonrugsa N., Toda T.;
RT   "Two kinesin-like Kin I family proteins in fission yeast regulate the
RT   establishment of metaphase and the onset of anaphase A.";
RL   Curr. Biol. 12:610-621(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Has a role in establishing metaphase during mitosis.
CC       Required for chromosome segregation where it generates tension
CC       during kinetochore capturing.
CC   -!- SUBUNIT: Heterodimer with klp6.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic microtubules in interphase,
CC       mitotic kinetochores in metaphase and spindle midzone in anaphase
CC       and telophase.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       II SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB072924; BAB69885.1; -.
DR   EMBL; Z97211; CAB10160.1; -.
DR   PIR; T40128; T40128.
DR   HSSP; P17119; 3KAR.
DR   GeneDB_SPombe; SPBC2F12.13; -.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Mitosis; Chromosome partition; Motor protein; Microtubules;
KW   ATP-binding; Coiled coil.
FT   DOMAIN        1    388       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      396    435       COILED COIL (POTENTIAL).
FT   DOMAIN      563    588       COILED COIL (POTENTIAL).
FT   NP_BIND     144    151       ATP (POTENTIAL).
SQ   SEQUENCE   883 AA;  99058 MW;  BB8A60E22DA397E8 CRC64;

  Query Match             6.6%;  Score 449;  DB 1;  Length 883;
  Best Local Similarity   21.3%;  Pred. No. 6.4e-13;
  Matches  218;  Conservative  159;  Mismatches  354;  Indels  294;  Gaps   37;

Qy          3 SKIKVVVRKRPLSELEKKK--KDSD--------------------IITVKNNCTLYIDE 39
Db          5 SSITVTVRVRPFSTAESANLIASSDRLSFGTSSSLRNPGSGRQIRRVVKVLDGRVLVFDP 64

Qy         40 PRYKVDMTKYIERH-----------------------EFIVDKVFDDTVDNFTVYENT 74
Db         65 PDETTATLSATNRRLSTSQQSLARLSRKSNNSAGFGRDLRYAFDRVFDETATQQQVYERT 124

Qy         75 IKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAAGDIFTFLNIY 134
Db        125 ARPLLDNILD-GFNATIFAYGATGCGKTHTISGTM-----QDPGLIYLTLKELFERMDHL 178

Qy        135 DKDNTKGIFISFYEIYCGKLYDLL-----QKRKMVAALENGKKEVVVKDLKILRVLTKEE 189
Db        179 RDEKIFDLRLSYLEIYNETIRDLLVSPTPNQAKPLNLREDADRRITVPGLTSLSPESLEE 238

Qy        190 LILKMIDGVLLRKIGVNSQNDESSRSHAILNIDL------KDINKNTSLGKIAFIDLAGS 243
Db        239 IIDIIMKGNANRTMSPTEANAASSRSHAVLQVTLIQKPRTAGINEDHTLATLSIIDLAGS 298

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAM--DSDKNHIPFRDSELTKVLRDIFVG 301
Db        299 ERATATKLRGSRL-FEGANINKSLLALGNCINALCDPHRRAHVPYRDSKLTRLLKFSLGG 357

Qy        302 KSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSE 361
Db        358 NCRTVMIVCVSPSSVHYEETHNTLKYANRAKNIKT---------EVLRNMISV------D 402

Qy        362 MNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYS 421
Db        403 RHVSQYVKAIVELREQISELENRLA-------QID----LSSQSNGSDQDAVTQSFAHES 451

Qy        422 SLNDIDKIKKNKKKGLINYKSTL-YNDNTINK----KHNNNNNNNNDNNDNNNDNNNNNN 476
Db        452 KLAEARNLLR------MTFEETLPLQNDTINKVEKVKH--------------------F 484

Qy        477 NDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLD 536
Db        485 DDSIRVLKYWL-------------------------SCYERILP--------------- 503

Qy        537 DMNCHLNNNDKSIFLHKKNLR--------------DNIKLKNRSSCDNIMNKKKNNLHL 581
Db        504 ------NSADERVFLVRSKLESLLTRRAEIIADIDPELVYQKFQRSVSHIINTYKQ---- 553

Qy        582 ARHSVGSKLTMFSYDPQKNKDNTFFKSNI--------NKMEDNTP--KDILYESRNVSNM 631
Db        554 ------EGATMYADVLQDEVD--LLKSIIENQVLDAQNKVDEFTPVLESLLRSSFKASSL 605

Qy        632 --NGNV----------LLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSN 679
Db        606 LKEGGMQELFSILEKWLLGIGLGEKPNISVLSESY---KLNS------TSDDSRTINRDR 656

Qy        680 MNS-----INNSNMNSNSIYKS---------------NYNSNQSISDV-QIRYVNEMDT 717
Db        657 VHSFPTQPLLNNNLPRMFFVKSPKKPVVFSKRSPKKRVRFDDSMSTSDSGASAYNSPIQT 716

Qy        718 SNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMK--- 774
Db        717 SKLKNMNFF-----------NTMHMPSTPAHKRPENKNQIDVE-INLTSPVSPMLEDKP 763
```

```
Qy        775 ---LYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNY---------------SMNFC 816
                 |   :     |  |:::|   |  :  :  | :: :              :||
Db        764 EPGLLIKSPLEKKQEVNSESTQLDQLLAEDSSTDDVSLPHLDTIDLDGSPVPKVPDLNFS 823

Qy        817 HYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGN 876
                  |::   ::  |||:      ||  | :  : |         :  |  ||:  : :|
Db        824 RANMDSPTFI--LNNEA-----IHNFDFSKPKTRQSLSSLTTLHLSNPANIIRKSLSMAE 876

Qy        877 NNSPR 881
              .  |   :
Db        877 NEEEK 881
```

```
RESULT 32
KF11_HUMAN
ID   KF11_HUMAN     STANDARD;      PRT;  1057 AA.
AC   P52732; Q15716;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin-like protein KIF11 (Kinesin-related motor protein Eg5)
DE   (Kinesin-like spindle protein HKSP) (Thyroid receptor interacting
DE   protein 5) (TRIP5) (Kinesin-like protein 1).
GN   KIF11 OR KNSL1 OR EG5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., PHOSPHORYLATION SITE THR-927, AND MUTAGENESIS.
RX   MEDLINE=96128120; PubMed=8548803;
RA   Blangy A., Lane H.A., D'Herin P., Harper M., Kress M., Nigg E.A.;
RT   "Phosphorylation by p34cdc2 regulates spindle association of human
RT   Eg5, a kinesin-related motor essential for bipolar spindle formation
RT   in vivo.";
RL   Cell 83:1159-1169(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98369052; PubMed=9701554;
RA   Whitehead C.M., Rattner J.B.;
RT   "Expanding the role of HsEg5 within the mitotic and post-mitotic
RT   phases of the cell cycle.";
RL   J. Cell Sci. 111:2551-2561(1998).
RN   [3]
RP   SEQUENCE OF 819-868 FROM N.A.
RX   MEDLINE=95295737; PubMed=7776974;
RA   Lee J.W., Choi H.-S., Gyuris J., Brent R., Moore D.D.;
RT   "Two classes of proteins dependent on either the presence or absence
RT   of thyroid hormone for interaction with the thyroid hormone
RT   receptor.";
RL   Mol. Endocrinol. 9:243-254(1995).
CC   -!- FUNCTION: MOTOR PROTEIN REQUIRED FOR ESTABLISHING A BIPOLAR
CC       SPINDLE. BLOCKING OF EG5 PREVENTS CENTROSOME MIGRATION AND ARREST
CC       CELLS IN MITOSIS WITH MONOASTRAL MICROTUBULE ARRAYS.
CC   -!- SUBUNIT: INTERACTS WITH THE THYROID HORMONE RECEPTOR IN THE
CC       PRESENCE OF THYROID HORMONE.
CC   -!- PTM: PHOSPHORYLATED EXCLUSIVELY ON SERINE DURING S PHASE, BUT ON
CC       BOTH SERINE AND THR-927 DURING MITOSIS, SO CONTROLLING THE
CC       ASSOCIATION OF EG5 WITH THE SPINDLE APPARATUS (PROBABLY DURING
CC       EARLY PROPHASE).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X85137; CAA59449.1; -.
DR   EMBL; U37426; AAA86132.1; -.
DR   EMBL; L40372; AAC41739.1; -.
DR   PIR; G02157; G02157.
DR   PDB; 1II6; 18-JUL-01.
DR   Genew; HGNC:6388; KIF11.
DR   GK; P52732; -.
DR   MIM; 148760; -.
DR   GO; GO:0005871; C:kinesin complex; TAS.
DR   GO; GO:0005819; C:spindle; TAS.
DR   GO; GO:0004002; F:adenosinetriphosphatase activity; TAS.
DR   GO; GO:0003777; F:microtubule motor activity; TAS.
DR   GO; GO:0007052; P:mitotic spindle assembly; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Mitosis;
KW   Phosphorylation; 3D-structure.
FT   DOMAIN       16    363       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      364    480       COILED COIL (POTENTIAL).
FT   DOMAIN      737    764       COILED COIL (POTENTIAL).
FT   NP_BIND     105    112       ATP (POTENTIAL).
FT   MOD_RES     927    927       PHOSPHORYLATION (BY CDC2).
FT   MUTAGEN     927    927       T->A: NO MITOTIC PHOSPHORYLATION. NO
FT                                BINDING TO SPINDLE APPARATUS.
FT   CONFLICT    674    676       RNS -> EL (IN REF. 2).
SQ   SEQUENCE   1057 AA;  119273 MW;  E322F2141BEF1601 CRC64;

  Query Match             6.5%;  Score 447.5;  DB 1;  Length 1057;
  Best Local Similarity   20.9%;  Pred. No. 9.1e-13;
  Matches  209;  Conservative  164;  Mismatches  387;  Indels  241;  Gaps   34;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVD-----MTKYIERHEFIVDK 59
              |:|||| ||  :   |:|       |:        |     |  | :|          :    |  :  |
Db         19 IQVVVRCRPFNLAERKASAHSIV----EC----DPVRKEVSVRTGGLADKSSRKTYTFDM 70

Qy         60 VFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLG----SQPYGQSD 115
              ||  :      || : :  | |:|    |  |: ||||||:|||:|| |      :: |   :
Db         71 VFGASTKQIDVYRSVVCP-ILDEVIMGYNCTIFAYGQTGTGKTFTMEGERSPNEEYTWEE 129

Qy        116 TP--GIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLL------QKRKMVAAL 167
               |  ||          ||  |   |         : :|  |||  :|:|||        :|   :
Db        130 DPLAGIIPRTLHQIFEKLT--DNGTEFSVKVSLLEIYNEELFDLLNPSSDVSERLQMFDD 187

Qy        168 ENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK--- 224
               |: |::| |: :  |   |:|:        |                 ||||:: :: :
Db        188 PRNKRGVIIKGLEEITVHNKDEVYQILEKGAAKRTTAATLMNAYSSRSHSVFSVTIHMKE 247
              .
Qy        225 ---DINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDK 281
                 |  :   ·:||: :|||||      : : :|: :   |  |||:|||  |    | |:
Db        248 TTIDGEELVKIGKLNLVDLAGSENIGRSGAVDKRAREAG-NINQSLLTLGRVITALVERT 306

Qy        282 NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCI 341
               |:|:|:|:||::|:|    |:::: :|| |||          |:||:||  |:  |  ||   ||
Db        307 PHVPYRESKLTRILQDSLGGRTRTSIIATISPASLNLEETLSTLEYAHRAKNILNK---- 362

Qy        342 NEEDDTNTERISILDSKGSEMNASSIENVVIK-----------------SNHLLSNNNN 383
                                     |:|      :  :||                    |  :  |
Db        363 -----------------PEVNQKLTKKALIKEYTEEIERLKRDLAAAREKNGVYISEEN 404

Qy        384 NKINRGKINDKIER--NNILKNKSFDKPREGFTSTFGKYSSLNDIDKIK---KNKKKGLI 438
              ::  ||:  :  |:    |    | |    :  ::       |      :  |::|:  |    :||  :  |
Db        405 FRVMSGKLTVQEEQIVELIEKIGAVEEELNRVTELF--MDNKNELDQCKSDLQNKTQELE 462

Qy        439 NYKSTLYN---------------DNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMV 483
                 :  |                  ::|  | |:  :      |     :       |  :  ::
Db        463 TTQKHLQETKLQLVKEEYITSALESTEEKLHDAASKLLNTVEETTKDVSGLHSKLDRKKA 522

Qy        484 NNMINHMINNNINNNIN--VNNNNNNNNNNSHNNHLPQPNYAFTD--TSDFSSLDDMN 539
              :  |    :       |:|    ||      : ::         |          |  :  :|  |:|| :
Db        523 VDQHNAEAQDIFGKNLNSLFNNMEELIKDGSSKQKAMLEVHKTLFGNLLSSSVSALDTIT 582
```

```
Qy        540 C----HLNNNDKSIFLHKKNL------RDNIKLKNRSSCDNIMNKKKNNL---------- 579
                   | :  :::  |    :         ::  ::::     ::|  | :|
Db        583 TVALGSLTSIPENVSTHVSQIFNMILKEQSLAAESKTVLQELINVLKTDLLSSLEMILSP 642

Qy        580 ---------HLARHSVGSKLTMF-SYDPQKNKDNTFFKS----NINKMEDNTPKDILYES 625
                       :|    : ||:     : || :::   |  |      |::::::||     ::
Db        643 TVVSILKINSQLKHIFKTSLTVADKIEDQKKRNSDGFLSILCNNLHELQENTICSLVESQ 702

Qy        626 RNVSNMNGNVLLGLNKNTHHD---------------ISTKDENHN------DNKINNGVI 664
               :     |:  :    |     |  ||                 :   |  ||          |
Db        703 KQCGNLTED--LKTIKQTHSQELCKLMNLWTERFCALEEKCENIQKPLSSVQENIQQKSK 760

Qy        665 NIINNSNVNS-------------INNSNM-------NSINNSNMNSNSIYKSNYNSNQSI 704
               :|:|     :|                :  |  |           |: :|:  :  :: | :  :  |
Db        761 DIVNKMTFHSQKFCADSDGFSQELRNFNQEGTKLVEESVKHSDKLNGNLEKISQETEQRC 820

Qy        705 SDVQIRYV----------NEMD----------------------------TSNKNNDNI 725
                :   |  |           || :                                :::    ||
Db        821 ESLNTRTVYFSEQWVSSLNEREQELHNLLEVVSQCCEASSSDITEKSDGRKAAHEKQHNI 880

Qy        726 FFDAISCDNNMYPNITNNNNNNN---------NNNNNNNNNIDVENYNNRDGTNNSMKLY 776
               | | :: |  | :      |   |    |              .|     :  :|:          ||     |  |
Db        881 FLDQMTIDEDKL--IAQNLELNETIKIGLTKLNCFLEQDLKLDIPT-----GTTPQRKSY 933

Qy        777 AYNS--------HNLFQPDNNKNTSNIQNINTNKNNQDGNV 809
               |  |               :|         |       :   :|  ::||::   :
Db        934 LYPSTLVRTEPREHLLDQLKRKQPELLMMLNCSENNKEETI 974
```

RESULT 33
KINH_MOUSE

```
ID   KINH_MOUSE     STANDARD;      PRT;   963 AA.
AC   Q61768; O08711; Q61580;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin heavy chain (Ubiquitous kinesin heavy chain) (UKHC).
GN   KIF5B OR KNS1 OR KHCS.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=97266654; PubMed=9112396;
RA   Meng Y.X., Wilson G.W., Avery M.C., Varden C.H., Balczon R.;
RT   "Suppression of the expression of a pancreatic beta-cell form of the
RT   kinesin heavy chain by antisense oligonucleotides inhibits insulin
RT   secretion from primary cultures of mouse beta-cells.";
RL   Endocrinology 138:1979-1987(1997).
RN   [2]
RP   SEQUENCE OF 1-881 FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=94224818; PubMed=8170981;
RA   Gudkov A.V., Kazarov A.R., Thimmapaya R., Axenovich S.A., Mazo I.A.,
RA   Roninson I.B.;
RT   "Cloning mammalian genes by expression selection of genetic
RT   suppressor elements: association of kinesin with drug resistance and
RT   cell immortalization.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:3744-3748(1994).
CC   -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC       PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT. MAY BE
CC       INVOLVED IN THE MECHANISMS OF GROWTH ARREST INDUCED BY EXPOSURE TO
CC       DNA-DAMAGING DRUGS OR BY CELLULAR SENESCENCE.
CC   -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC       CHAINS.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC       TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC       KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC       ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC       DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC       INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC       VESICLES AND MEMBRANOUS ORGANELLES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       SUBFAMILY.
CC   
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U86090; AAB53940.1; -.
DR   EMBL; L27153; AAA20133.1; -.
DR   EMBL; L29223; AAA20132.2; -.
DR   HSSP; P33176; 1BG2.
DR   MGD; MGI:1098268; Kif5b.
DR   GO; GO:0005515; F:protein binding activity; IPI.
DR   GO; GO:0007028; P:cytoplasm organization and biogenesis; IMP.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    328       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      329    914       COILED COIL.
FT   DOMAIN      915    963       GLOBULAR.
FT   NP_BIND      85     92       ATP (BY SIMILARITY).
FT   CONFLICT     20     21       ES -> AT (IN REF. 2).
FT   CONFLICT     40     40       V -> M (IN REF. 2).
FT   CONFLICT     75     75       E -> G (IN REF. 2).
FT   CONFLICT    191    191       H -> D (IN REF. 2).
FT   CONFLICT    255    255       N -> K (IN REF. 2).
FT   CONFLICT    396    396       I -> T (IN REF. 2).
FT   CONFLICT    403    403       G -> P (IN REF. 2).
FT   CONFLICT    451    451       T -> I (IN REF. 2; AAA20132).
FT   CONFLICT    520    520       T -> S (IN REF. 2).
FT   CONFLICT    523    523       F -> L (IN REF. 2).
FT   CONFLICT    653    653       D -> V (IN REF. 2).
FT   CONFLICT    664    664       L -> V (IN REF. 2).
FT   CONFLICT    744    744       T -> Q (IN REF. 2).
FT   CONFLICT    814    814       S -> R (IN REF. 2).
FT   CONFLICT    865    865       F -> K (IN REF. 2).
FT   CONFLICT    881    881       L -> R (IN REF. 2).
SQ   SEQUENCE   963 AA;  109549 MW;  718CDBD102E0475E CRC64;

  Query Match             6.5%;  Score 447;  DB 1;  Length 963;
  Best Local Similarity   21.3%;  Pred. No. 8.6e-13;
  Matches  211;  Conservative  177;  Mismatches  414;  Indels  188;  Gaps   37;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 64
                |||: | |||:| |   :  |                |:   ::: :  |   |     :   |:||   :
Db          9 IKVMCRFRPLNESEVNRGDK-----------YV--AKFQGEDTVVIASKPYAFDRVFQSS 55

Qy         65 VDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGS--QPYGQSDTPGIF 120
                       || :   | :: |: |   || :    ||||| ||||:|| |     ||      | |
Db         56 TSQEQVYNDCAKKIVKDVLEGYNGTI---FAYGQTSSGKTHTMEGKLHDPEGMGIIPRIV 112

Qy        121 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEV-VVKDL 179
                   |      ||| ::     |::     | :|::|||   |: |||     |       ::    |    |    ||
Db        113 Q----DIFNYIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLSVHEDKNRVPYVKGC 168

Qy        180 KILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTS---LGKIA 236
                  |  : :|::   : :|     |   :  |   :  |: |||||:|    |::|    |     |        ||:
Db        169 TERFVCSPDEVMDTIDEGKSNRHVAVTNMNEHSSRSHSIFLINVKQENTQTEQKLSGKLY 228

Qy        237 FIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLR 296
                :||||||: :  | ::          :    |||:|| ||       |  |:          ::|:|||::|::|:
Db        229 LVDLAGSEKVSKTGAEG-AVLDEAKNINKSLSALGNVISALAEGSTYVPYRDSKMTRILQ 287
```

```
Qy        297 DIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILD 356
Db        288 DSLGGNCRTTIVICCSPSSYNESETKSTLLFGQRAKTIKN-TVCVNVE--LTAEQWKKKY 344

Qy        357 SKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPR---EGF 413
Db        345 EKEKEKN-KTLRNTI--------QWLENELNRWRNGETVP-----IDEQFDKEKANLEAF 390

Qy        414 T------------STFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNN 460
Db        391 TADKDIAITSDKGAAAVGMAGSFTDAERRKCEEELAKL-YKQL---------------- 432

Qy        461 NNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLP 520
Db        433 ---------DDKDEEINQQSQLVEKLKTQMLDQE-----ELLASTRRDQDNMQAELNRLQ 478

Qy        521 QPNYAFTD--TSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNN 578
Db        479 AENDASKEEVKEVLQALEEL--AVNYDQKS-----QEVED--KTKEYELLTDEFNQKSAT 529

Qy        579 LHLARHSVGSKLTMFSYDPQKNKDNT-FFKSNINKMEDNTPKDI--------------- 621
Db        530 L----ASIDAEL-------QKLKEMTNHQKKRAAEMMASLLKDLAEIGIAVGNNDVKQPE 578

Qy        622 ------------LYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIIN 668
Db        579 GTGMIDEEFTVARLYISKMKSEVKTMVKRCKQLESTQTESNKKMEENEKELAACQLRI-- 636

Qy        669 NSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISD--VQIRYVNEMDTSNKNNDNIF 726
Db        637 -----SQHEAKIKSLTEYLQNDEQKKRQLEESLDSLGEELVQLRAQEKVHEMEKEHLNKV 691

Qy        727 FDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYN-----SH 781
Db        692 QTANEVKQAVEQQIQSHRETHQKQISSLRDEVEAKEKLITDLQDQNQKMVLETERLRVEH 751

Qy        782 NLFQPDNNKNTSNIQNINTNKNNQ---------DGNVNYSMNFCHYNLNDKNYLIDLNN 831
Db        752 ERLKATDQEKSRKLHELTVMQDRREQARQDLKGLEETVAKELQTLH-NLR-KLFVQDLAT 809

Qy        832 KEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNN--NIVIVNNNMGNNNS------PRMK 883
Db        810 RVKKSAEVDSDDTG-----GSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLE 864

Qy        884 YGLCGSHTSIDNMKNNEMKNNEMKDNEMKD 913
Db        865 FRL---RATAERVKALESALKEAKENASRD 891
```

```
RESULT 34
KF14_HUMAN
ID   KF14_HUMAN     STANDARD;      PRT;  1648 AA.
AC   Q15058;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein KIF14.
GN   KIF14 OR KIAA0042.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96051398; PubMed=7584044;
RA   Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA   Seki N., Kawarabayasi Y., Ishikawa K.-I., Tabata S.;
RT   "Prediction of the coding sequences of unidentified human genes. II.
RT   The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 1:223-229(1994).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D26361; BAA05392.1; -.
DR   HSSP; P17119; 3KAR.
DR   Genew; HGNC:19181; KIF14.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN      355    628       KINESIN-MOTOR.
FT   DOMAIN      705    791       COILED COIL (POTENTIAL).
FT   DOMAIN      825    891       FHA.
FT   DOMAIN      922   1079       COILED COIL (POTENTIAL).
FT   DOMAIN     1332   1348       COILED COIL (POTENTIAL).
FT   DOMAIN     1468   1500       COILED COIL (POTENTIAL).
FT   NP_BIND     447    454       ATP (POTENTIAL).
SQ   SEQUENCE   1648 AA;  186491 MW;  FB1423668A7B79D7 CRC64;

  Query Match             6.5%;  Score 446;  DB 1;  Length 1648;
  Best Local Similarity   19.8%;  Pred. No. 1.7e-12;
  Matches  284;  Conservative  213;  Mismatches  507;  Indels  428;  Gaps   59;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDI-------ITVKNNCTLYIDEPRYKVDMTKYIERHE 54
Db        356 NSQVTVAVRVRPFTKREKIEKASQVVFMSGKEITVEHPDTKQVYNFIYDVSFWSFDECHP 415

Qy         55 FIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQS 114
Db        416 HYASQT--------TVYEKLAAPLLERAFEGFNTC-LFAYGQTGSGKSYTMMGF-----S 461

Qy        115 DTPGIFQYAAGDIFT-FLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKE 173
Db        462 EEPGIIPRFCEDLFSQVARKQTQEVSYHIEMSFFEVYNEKIHDLL-----VCKDENGQRK 516

Qy        174 VVVKDLKILRVLTK-------EELILKMID-----------GVLLRKIGVNSQNDESSRS 215
Db        517 ------QPLRVREHPVYGPYVEALSMNIVSSYADIQSWLELGNKQRATAATGMNDKSSRS 570

Qy        216 HAILNIDLKDINKNTSLG---------KIAFIDLAGSERGADTVSQNKQTQTDGANINRS 266
Db        571 HSVFTLVMTQTKTEFVEGEEHDHRITSRINLIDLAGSER-CSTAHTNGDRLKEGVSINKS 629

Qy        267 LLALKECIRAMDSDKNH----IPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTL 322
Db        630 LLTLGKVISALSEQANQRSVFIPYRESVLTWLLKESLGGNSKTAMIATISPAASNIEETL 689

Qy        323 NTLRYSSRVKNFKNKSTCINEEDDTNTERI-----SILDSKGSEMNASSIENVVIKSNHL 377
Db        690 STLRYANQARLIVNIAK-VNE--DMNAKLIRELKAEIAKLKAAQRNSRNID----PERYR 742

Qy        378 LSNNNNNKINRGKIN----DKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNK 433
Db        743 LCRQEITSL-RMKLHQQERDMAEMQRVWKEK-FEQAEK-----------RKLQETKELQ 788

Qy        434 KKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINN 493
Db        789 KAGIM-----------------------FQMDNHLPNLVNLNED--PQLSEMLLYMIKE 822

Qy        494 NINNNINVNNNNNNNNNNNNSH----------NNHLPQPNYAFTDTSDFSSLDDMNCHLN 543
```

```
                                 |:||           ::|    |  |  |     : :   ::|
Db        823 --------GTTTVGKYKPNSSHDIQLSGVLIADDHCTIKN--FGGTVSIIPVGEAKTYVN 872

Qy        544 -----------NNDKSI--------FLHKKNLRDNIKLKNRSSCDNIMNK-------KKN 577
                          : |: |         | |  ::  | |  |  |  :::         ||
Db        873 GKHILEITVLRHGDRVILGGDHYFRFNHPVEVQ---KGKRPSGRDTPISEGPKDFEFAKN 929

Qy        578 NLHLARHSVGSKLTMFSYDPQ-KNKDNTFFKSNINK---MEDNTPKDILYESRNVSNMNG 633
                | :|:     :|     : | | | |:        | |      :: : :   |||:
Db        930 ELLMAQR---SQLEAEIKEAQLKAKEEMMQGIQIAKEMAQQELSSQKAAYESK------- 979

Qy        634 NVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSI 693
                      |    ::  : :     :|||      |  :  :  :  :     :|   :
Db        980 ------IKALEAELREESQRKKMQEINNQKA----NHKIEELEKAKQHLEQEIYVNKKRL 1029

Qy        694 YKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNN 753
                      : |:: |  ||:   ::          |  :
Db       1030 EMETLATKQALEDHSIRHARILEALETEKQKIAKEV----------------------- 1065

Qy        754 NNNNIDVENYNNRDGTN------NSMKL-----------------YAYNSHNLFQPDNNK 790
                    |  :| |||| |           :||||                | :  |::    ::
Db       1066 ---QILQQNRNNRDKTFTVQTTWSSMKLSMMIQEANAISSKLKTYYVFGRHDISDKSSSD 1122

Qy        791 NTSNIQNI-----------------------NTNKNNQDGNVNYSMNFCHYNLNDKNYL 826
              :  ::|:                                :| :|:  :   ||      :  :
Db       1123 TSIRVRNLKLGISTFWSLEKFESKLAAMKELYESNGSNRGEDA-----FCDPEDEWEPDI 1177

Qy        827 IDL---------NNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNN 877
              |           :    |:: | || ::|              |  :  ::: ::  :: : :|
Db       1178 TDAPVSSLSRRRSRSLMKNRRISGCLHDI-----QVHPIKNLHSSHSSGLMDKSSTIYSN 1232

Qy        878 NSPRMKYGLC----GS------------HTSIDNMKNNEM----------KNNEMKDNEM 911
               ::      |:|    ||              | |:: |: :               | :| :| |
Db       1233 SAESFLPGICKELIGSSLDFFGQSYDEERTIADSLINSFLKIYNGLFAISKAHEEQDEES 1292

Qy        912 KDNHIKSNNNNSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNI--NSNIY 969
              :||   |:   |                    | |   |   :     :  :   :::   :||
Db       1293 QDNLFSSDRAIQSL-----------------TIQT-ACAFEQLVVLMKHWLSDLLPCTNIA 1335

Qy        970 Q-NDDIIYTINSLNDY---------------------------------------MSNTL 989
              :   |::   :  |                                          :  :
Db       1336 RLEDELRQEVKKLGGYLQLFLQGCCLDISSMIKEAQKNAIQIVQQAVKYVGQLAVLKGSK 1395

Qy        990 LHFKEKYTYPTLSTNEDIYNKEME----GKHIRLDD-QDKYDDNDNNNVDNNNKNNVDNN 1044
              ||| |         | |:  :   :       |  |  ||  :      :     ||  |
Db       1396 LHFLENGNNKAASVQEEFMDAVCDGVGLGMKILLDSGLEKAKELQHELFRQCTKNEVTKE 1455

Qy       1045 VDNNNVD-----NNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNN--EYLSYFQKNVDT 1097
              :  | :           |:   |         |    ::: :: :  |    |   |:|
Db       1456 MKTNAMGLIRSLENIFAESKIKSFRRQV--QEENFEYQDFKRMVNRAPEFLK-------- 1505

Qy       1098 IINNCLNSLDISSMYDDTKEILNNILLSKYK-AEKDNVIKKYINEDIKNMSLE-----EI 1151
               : :||              |     |::|  |     |    :  :     | |:|:: :    :
Db       1506 -LKHCL-------------EKAIEIIISALKGCHSDINLLQTCVESIRNLASDFYSDFSV 1551

Qy       1152 DKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPD 1203
                |:  ||  |                    |     :: |    |'| |: |      :||
Db       1552 PSTSVGSYESR--------------VTHIVHQELESLAKSLLFCFESEESPD 1589
```

RESULT 35
KF3A_MOUSE

```
ID   KF3A_MOUSE     STANDARD;      PRT;   701 AA.
AC   P28741;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Kinesin-like protein KIF3A (Microtubule plus end-directed kinesin
DE   motor 3A).
GN   KIF3A OR KIF3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93077686; PubMed=1447303;
RA   Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M., Hirokawa N.;
RT   "Kinesin family in murine central nervous system.";
RL   J. Cell Biol. 119:1287-1296(1992).
CC   -!- FUNCTION: MICROTUBULE-BASED ANTEROGRADE TRANSLOCATOR FOR
CC       MEMBRANOUS ORGANELLES. PLUS END-DIRECTED MICROTUBULE SLIDING
CC       ACTIVITY IN VITRO.
CC   -!- SUBUNIT: HETERODIMER OF KIF3A AND KIF3B.
CC   -!- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN
CC       TISSUE (MAINLY IN THE CEREBELLAR GRANULAR LAYER) WITHIN A SINGLE
CC       TYPE OF NEURONAL CELL.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       II SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D12645; BAA02166.1; -.
DR   PIR; B44259; B44259.
DR   HSSP; P17119; 3KAR.
DR   MGD; MGI:107689; Kif3a.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Neurone.
FT   DOMAIN        1    350       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      351    586       COILED COIL (BY SIMILARITY).
FT   DOMAIN      587    701       GLOBULAR.
FT   NP_BIND     100    107       ATP (BY SIMILARITY).
FT   DOMAIN      442    445       POLY-GLU.
FT   DOMAIN      509    512       POLY-ARG.
SQ   SEQUENCE   701 AA;  80167 MW;  2405872DF2D85A29 CRC64;

  Query Match             6.5%;  Score 445.5;  DB 1;  Length 701;
  Best Local Similarity   29.6%;  Pred. No. 7.1e-13;
  Matches  145;  Conservative   84;  Mismatches  196;  Indels   65;  Gaps   16;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHE----FIVDKV 60
              :|||||  |||:| ||                 ::|     |  |  :  |  :|    |  |  |
Db         15 VKVVVRCRPLNEREKSMCYRQAVSV--------DEMRGTITVHKTDSSNEPPKTFTFDTV 66

Qy         61 FDDTVDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPG 118
                |         ||  |  :|:|  :  |  || :   ||||||:|||:|| |  :          |
Db         67 FGPESKQLDVYNLTARPIIDSVLEGYNGTI---FAYGQTGTGKTFTMEGVR--AVPGLRG 121

Qy        119 IFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQK--RKMVAALENGKKEVVV 176
               :   :    ||  :    :  |     : :|: |||  :: |||  |     :  :     |     |  :
Db        122 VIPNSFAHIFGHIAKAEGDTRFLVRVSYLEIYNEEVRDLLGKDQTQRLEVKERPDVGVYI 181

Qy        177 KDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK------DINKNT 230
               |||    |     :::     |   |     | :|   : |:  ||||||   |  ::           | |  :
Db        182 KDLSAYVVNNADDMDRIMTLGHKNRSVGATNMNEHSSRSHAIFTITIECSEKGVDGNMHV 241

Qy        231 SLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM-DSDKNHIPFRDS 289
               :||:   :||||||  |  |  :  :  :  :  :     ||  ||    |     |  |:  |      |:|:|:|
Db        242 RMGKLHLVDLAGSERQAKTGATGQRLK-EATKINLSLSTLGNVISALVDGKSTHVPYRNS 300

Qy        290 ELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEED---- 345
               :||::|:|    | ||::| |||  |          ::|::||||::| || |||:   |||:
Db        301 KLTRLLQDSLGGNSKTMMCANIGPADYNYDETISTLRYANRAKNIKNKAR-INEDPKDAL 359
```

```
Qy        346 ----DTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNIL 401
                  |  :     :| |:: | |                  :  |  | |  |: :  :
Db        360 LRQFQKEIEELKKKLEEGEEVSGSDISGSEEDDEEGELGEDGEK--RKKRRDQAGKKKVS 417

Qy        402 KNK------SFDKPREGFTSTFGKYSSLNDIDKIKKNK--------KKGLI----NYKST 443
               :|         |: |:    :           |::: ::||        :| |:      ::|
Db        418 PDKMVEMQAKIDEERKALETKL-------DMEEEERNKARAELERREKDLLKAQQEHQSL 470

Qy        444 LYNDNTINKK 453
              |   : : ||
Db        471 LEKLSALEKK 480


RESULT 36
KF5C_MOUSE
ID   KF5C_MOUSE     STANDARD;      PRT;   956 AA.
AC   P28738; Q9Z2F8;
DT   01-DEC-1992 (Rel. 24, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin heavy chain isoform 5C (Kinesin heavy chain neuron-specific
DE   2).
GN   KIF5C OR NKHC2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   PRELIMINARY SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Brain;
RA   Kato K.;
RT   "A collection of cDNA clones with specific expression patterns in
RT   mouse brain.";
RL   Eur. J. Neurosci. 2:704-711(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99000842; PubMed=9782088;
RA   Xia C., Rahman A., Yang Z., Goldstein L.S.B.;
RT   "Chromosomal localization reveals three kinesin heavy chain genes in
RT   mouse.";
RL   Genomics 52:209-213(1998).
CC   -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC       PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC   -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC       CHAINS.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC       TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC       KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC       ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC       DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC       INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC       VESICLES AND MEMBRANOUS ORGANELLES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       SUBFAMILY.
CC   -!- CAUTION: REF.1 SEQUENCE SEEMS TO HAVE BEEN A HYBRID OF A KIF5A
CC       AND A KIF5C SEQUENCE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X61435; CAA43677.1; -.
DR   EMBL; AF067180; AAC79804.1; -.
DR   HSSP; P56536; 2KIN.
DR   MGD; MGI:1098269; Kif5c.
DR   GO; GO:0005515; F:protein binding activity; IPI.
DR   GO; GO:0008045; P:motor axon guidance; IMP.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    386       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      406    923       COILED COIL.
FT   DOMAIN      859    956       GLOBULAR.
FT   DOMAIN      174    315       MICROTUBULE-BINDING.
FT   NP_BIND      86     93       ATP (BY SIMILARITY).
SQ   SEQUENCE   956 AA;  109240 MW;  D5A8C701A2911AB2 CRC64;

  Query Match             6.5%;  Score 443.5;  DB 1;  Length 956;
  Best Local Similarity   21.8%;  Pred. No. 1.2e-12;
  Matches  216;  Conservative  162;  Mismatches  426;  Indels  185;  Gaps   38;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYI-ERHEFIVDKVFDD 63
              |||: | |||:| |  : |  |             |::| : |  |  :   :: |:|
Db          9 IKVMCRFRPLNEAEILRGDKFI------------PKFKGEETVVIGQGKPYVFDRVLPP 55

Qy         64 TVDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQ 121
                  ||    | :: |: |  || :   ||||| ||||:|| |              ||
Db         56 NTTQEQVYNACAKQIVKDVLEGYNGTI---FAYGQTSSGKTHTMEGK--LHDPQLMGIIP 110

Qy        122 YAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEV-VVKDLK 180
                | |||  :    |::    | :|::|||  |: |||   |    |:    |  |  ||
Db        111 RIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCT 170

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTS---LGKIAF 237
                  | : ||::   : :|    | : | : | :|: |||||:|   |::|   |  |     ||:
Db        171 ERFVSSPEEVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYL 230

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM-DSDKNHIPFRDSELTKVLR 296
              :|||||: : | .::        :  |||:|| ||   | |: :   | |:|:|||::|:|:
Db        231 VDLAGSEKVSKTGAEG-AVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQ 289

Qy        297 DIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCIN------------E 343
              . |    | :: ::      ||::    :| :|| :   |  |  || :  :|           :
Db        290 DSLGGNCRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKN-TVSVNLELTAEEWKKKYEK 348

Qy        344 EDDTNTERISILDSKGSEMN-------ASSIENVVIKSNHLLSNNNNNKI--NRGKINDK 394
               | : |     |:|      |:|            | :   | :  |    :|  |   |   : |
Db        349 EKEKNKALKSVLQHLEMELNRWRNGEAVPEDEQISAKDHKSLEPCDNTPIIDNITPVVDG 408

Qy        395 IERNNILKNKSFDKPREGFTSTF----GKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTI 450
              |              :|  |  ||  :       |    :|  . :: :  |: :::  :
Db        409 ISAEK-------EKYDEEITSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRD 461

Qy        451 NKKHNNNNNNNNDNNDNNNDNNNN----------NNNDSSSMVNN--MINHMINNNINNN 498
               :|         |:   |                    |  :   |   :   :  :  :
Db        462 YEKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQK 521

Qy        499 INVNNNNNNNNNNNSHNNHLPQP-----NYAFTD---------TSDFSSLDDMNCHLNN 544
                    :      :|| :        |       |          |:|  :| |:|   :
Db        522 TTTLTTTQRELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEE 581

Qy        545 NDKSIFLHKKNLRDNIK-LKNRS----SCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQK 599
                  |:   :: :| | |||     |          |:| |    :     :   :. |    |
Db        582 EFTMARLYISKMKSEVKSLVNRSKQLESAQMDSNRKMN---ASERELAACQLLISQHEAK 638

Qy        600 NKDNTFFKSNI----NKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHN 655
                | |  :  |:      ::|::  :| |  |    :: :                :   |::| :|
Db        639 IKSLTDYMQNMEQKRRQLEES--QDSL--SEELAKLRA-------QEKMHEVSFQD---- 683

Qy        656 DNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVN-- 713
                    |   :  :  :   |          ||  ::      |    :        | | | :|| :|
Db        684 --KEKEHLTRLQDAEEVKKALEQQMESHREAHQKQLSRLRDEIEEKQRIID-EIRDLNQK 740

Qy        714 -EMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNS 772
               :::      ::|                                     | : :|:
Db        741 LQLEQERLSSD------------------------------YNKLKIEDQEREVKLEKL 769
```

```
Qy         773 MKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNK 832
                 : |           |:|      :     :    |:  :     | ||  | :: ||   :
Db         770 LLL-----------NDKR----EQAREDLKGLEETVSIELQTLH-NLR-KLFVQDLTTR 811

Qy         833 EQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNN--NIVIVNNNMGNNNS------PRMKY 884
                 :|    :     |          :|:| :|  ||     :   |:   :  :|:         |:::
Db         812 VKKSVELDSDDGG-----GSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEK 866

Qy         885 GLCGSHTSIDNMKNNEMKNNEMKDNEMKD 913
                 |      : : :|  |     | |:| |:|
Db         867 RL---RATAERVKALESALKEAKENAMRD 892


RESULT 37
KF4A_XENLA
ID   KF4A_XENLA     STANDARD;      PRT;  1226 AA.
AC   Q91784; Q9PSI0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Chromosome-associated kinesin KLP1 (Chromokinesin).
GN   KLP1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP   SPECIFICITY.
RC   TISSUE=Oocyte;
RX   MEDLINE=95236444; PubMed=7720067;
RA   Vernos I., Raats J., Hirano T., Heasman J., Karsenti E., Wylie C.;
RT   "Xklp1, a chromosomal Xenopus kinesin-like protein essential for
RT   spindle organization and chromosome positioning.";
RL   Cell 81:117-127(1995).
RN   [2]
RP   SEQUENCE OF 9-338 FROM N.A.
RX   MEDLINE=93246065; PubMed=8482413;
RA   Vernos I., Heasman J., Wylie C.;
RT   "Multiple kinesin-like transcripts in Xenopus oocytes.";
RL   Dev. Biol. 157:232-239(1993).
CC   -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
CC       SPINDLE STABILIZATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR, ASSOCIATED WITH MITOTIC
CC       CHROMOSOMES.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN OOCYTES, EGGS, TESTES AND
CC       BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC       CHROMOKINESIN SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X82012; CAA57539.1; -.
DR   PIR; I51617; I51617.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; DNA-binding;
KW   Nuclear protein; Coiled coil.
FT   DOMAIN        1    350       KINESIN-MOTOR.
FT   DOMAIN      351   1006       COILED COIL (BY SIMILARITY).
FT   DOMAIN     1007   1226       GLOBULAR.
FT   NP_BIND      87     94       ATP (POTENTIAL).
FT   CONFLICT    163    163       I -> L (IN REF. 2).
SQ   SEQUENCE   1226 AA;  138923 MW;  7F0275FCF3316697 CRC64;

  Query Match             6.5%;  Score 442.5;  DB 1;  Length 1226;
  Best Local Similarity   24.9%;  Pred. No. 1.8e-12;
  Matches  174;  Conservative  118;  Mismatches  275;  Indels  131;  Gaps   24;

Qy           5 IKVVVRKRPLSELEKKKKDSDIITVKNN-----CTLYIDEPRYKVDMTKYIERHEFIVDK 59
                 ::| :| |||                  : :||      |  ::    :   :  |:           |  |
Db           9 VRVALRCRPL------------VPKENNEGCKMCLTFVPGEQQVIVGTE----KSFTYDY 52

Qy          60 VFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGI 119
                 ||| : :    || : : |||   |:: |     :   ||||||||||:| |:  :  | : | :
Db          53 VFDPSAEQEEVYNSAVAPLIKGLFK-GYNATVLAYGQTGSGKTYSMGGAYTHNQENEPTV 111

Qy         120 FQYAAGDIFTFLNIYDKDNTK-GIFISFYEIYCGKLYDLL----QKRKMVAALENGKKEV 174
                          |   |  |: :    :  : :|: |||   :: |||       |    ::   |: |: :
Db         112 GVIPRTVIALFREIHQRPEWEFNLKVSYLEIYNEEILDLLYAARDKTNTISIREDPKEGI 171

Qy         175 VVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDL---KDINKNTS 231
                 :   |          | |   : :   :   |     :     :  | :|||||||    | :     |: :|| |
Db         172 KICGLTERDVKTALDTLSCLEQGNSSRTVASTAMNSQSSRSHAIFTISIEQRKEGDKNNS 231

Qy         232 L-GKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM-DSDK--NHIPFR 287
                      |:   :||||||     | ::   :  : :| :||| ||  |      | |:  |     |      :|:|
Db         232 FRSKLHLVDLAGSERQKKTKAEGDRLK-EGISINRGLLCLGNVISALGDESKKGGFVPYR 290

Qy         288 DSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDT 347
                 ||:||::|:|      |  |  | ::|||  :||    |      |:||||||||: |  :     |||              :
Db         291 DSKLTRLLQDSLGGNSHTLMIACVSPADSNMEETLNTLRYADRARKIKNKPIVNTDPQAA 350

Qy         348 NTERIS---------ILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERN 398
                      :|:             :|  :  |    :       :  |  :     | :|  |          ||  |     |  | |: |
Db         351 ELQRLKLQVQELQVLLLQAHGGTL---PVLNSMEPSENLQSLMERNK-NLEKENGKLSRE 406

Qy         399 NILKNKSFDKPREGFTSTFGKYSSLNDIDKI----KKNKKKGLINYKSTLYNDNTINKKH 454
                      |     :               |: |              ::||      ::|:|     :       :        :   :| :
Db         407 --LGEAAVQ------TAQF--------LEKIIMTEQQNEKLGSKMEELKQHAACKVNLQR 450

Qy         455 NNNNNNNNDNNDNNNDNNNNNN-----NDSSSMVNNMINHMINNNINNNINVNNNNNNNN 509
                           :  :   ||     |                    |  ||  :      |       |           :     :    ::     :
Db         451 LVETLEDQELKDNVEVIQNLQQVIVQLQDESSGIAGSIEAMDEEAASFPVPEEDSGEKRS 510

Qy         510 NNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHL----------NNNDKSIFLHKKNLRDN 559
                 ::   :  |: |  |      :                      | ::|    |                  ||:  :     :      :|
Db         511 SDGFTTNHALRQAQLS-------KELIELNKALVMKEALAKKMAQNDRQLEPIQSEYLNN 563

Qy         560 IKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFK------------ 607
                 |    |:    |       ::  |:|    |  || |  ||                          |||    |
Db         564 I--KHLESEVGVLQKEKEELILALHSA-------------KKDNNQAKLSERRRKRLQEL 608

Qy         608 --------------SNINKMEDNTPKDILYESRNVSNM 631
                                    | :  |:  ::|  |  :        :: :     |
Db         609 EGQMTELKKKLGEQSKLLKLRESTEKTVAKMNQEIQGM 646


RESULT 38
EG52_XENLA
ID   EG52_XENLA     STANDARD;      PRT;  1067 AA.
AC   Q91783;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin-related motor protein Eg5 2.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Egg;
```

```
RX   MEDLINE=94299012; PubMed=8026619;
RA   Houliston E., le Guellec R., Kress M., Philippe M., le Guellec K.;
RT   "The kinesin-related protein Eg5 associates with both interphase and
RT   spindle microtubules during Xenopus early development.";
RL   Dev. Biol. 164:147-159(1994).
CC   -!- FUNCTION: PLUS END-DIRECTED MOTOR PROTEIN REQUIRED FOR
CC       ESTABLISHING A BIPOLAR SPINDLE. ASSOCIATES WITH BOTH INTERPHASE
CC       AND SPINDLE MICROTUBULES. MAY BE INVOLVED IN NUCLEAR DIVISIONS
CC       TAKING PLACE DURING THE DEVELOPMENT OF UNFERTILIZED EGGS.
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO HEAVY AND TWO LIGHT CHAINS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CONCENTRATED AROUND THE POLAR ENDS OF
CC       BOTH MEIOTIC AND MITOTIC SPINDLES.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN UNFERTILIZED EGGS,
CC       ESPECIALLY IN THE GERMINAL VESICLE AND IN THE RADIAL YOLK-POOR
CC       CHANNELS. ALSO PRESENT IN TESTIS.
CC   -!- DEVELOPMENTAL STAGE: ACCUMULATES DURING THE LATTER STAGES OF
CC       OOGENESIS AND LEVELS INCREASE THREE-FOLD DURING OOCYTE MATURATION.
CC       LEVELS DECREASE AFTER FERTILIZATION.
CC   -!- PTM: PHOSPHORYLATION DURING MITOSIS AT THR-937 CONTROLS THE
CC       ASSOCIATION OF EG5 WITH THE SPINDLE APPARATUS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC       SUBFAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X71864; CAA50695.1; -.
DR   PIR; I51616; S33417.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Mitosis;
KW   Phosphorylation.
FT   DOMAIN       16    363       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      364    487       COILED COIL (POTENTIAL).
FT   DOMAIN      692    736       COILED COIL (POTENTIAL).
FT   DOMAIN      884    920       COILED COIL (POTENTIAL).
FT   DOMAIN     1040   1067       COILED COIL (POTENTIAL).
FT   NP_BIND     105    112       ATP (BY SIMILARITY).
FT   MOD_RES     937    937       PHOSPHORYLATION (BY CDC2) (BY
FT                                SIMILARITY).
SQ   SEQUENCE   1067 AA;  119885 MW;  48A477D1E2559115 CRC64;

  Query Match             6.4%;  Score 439.5;  DB 1;  Length 1067;
  Best Local Similarity   22.1%;  Pred. No. 2e-12;
  Matches  198;  Conservative  156;  Mismatches  327;  Indels  213;  Gaps   31;

Qy          1 MNSK-------IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKV-----DMTK 48
                |:||        |:|||| || ::||:|             ::  |  :  | :|     ::
Db          8 MSSKKDDKGKNIQVVVRCRPFNQLERK--------ASSHSVLECESQRKEVCVRTGEVND 59

Qy         49 YIERHEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS 108
                : :   :   | ||           || : : |  |:|      |  |: |||||||:|||:|| |
Db         60 KLGKKTYTFDMVFGPAAKQIDVYRSVVCP-ILDEVIMGYNCTIFAYGQTGTGKTFTMEGE 118

Qy        109 Q------PYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQ--- 159
                :       :  |        ||         ||  |:  :     :  :|  |||  :|:|||
Db        119 RSSDEEFTWEQDPLAGIIPRTLHQIFEKLS--EIGTEFSVKVSLLEIYNEELFDLLSPSP 176

Qy        160 ---KRKMVAALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSH 216
                 :|   :  | :::| |: : |  |:|:    :    ||      |  |||||
Db        177 DVGERLQMFDDPRNKRGVIIKGLEEISVHNKDEVYQILERGAAKRKTASTLMNAYSSRSH 236

Qy        217 AILNIDLK------DINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLAL 270
                :: :: :         |  :   :||:  :|||||     : : :|: :  | |||:||| |
Db        237 SVFSVTIHMKETTIDGEELVKIGKLNLVDLAGSENIGRSGAVDKRAREAG-NINQSLLTL 295

Qy        271 KECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSR 330
                   | |:         |||:|:|:|:||::|:|    |::|: :|| :||      |:|::|| |:||
Db        296 GRVITALVERAPHIPYRESKLTRILQDSLGGRTKTSIIATVSPASINLEETMSTLDYASR 355

Qy        331 VKNFKNKSTCINEEDDTNTERISILD-SKGSEMNASSIENVVIKSNHLLSNNNNNKINRG 389
                ||  ||       |  :     |::  | : ::         :       |:     ||| |   :: :|
Db        356 AKNIMNKP----EVNQKLTKKALIKEYTEEIERLKRELATAREKNGVYLSNENYEQL-QG 410

Qy        390 KI-----------------NDKIER----------------------------NNILK 402
                |:                  ::|:|                            ||: :
Db        411 KVLSQEEMITEYSEKIAAMEEEIKRIGELFADNKKELEECTTILQCKEKELEATQNNLQE 470

Qy        403 NKSFDKPREGFTSTFGKYSSLNDID-KIKKNKKKGLINYKSTLYNDNTINKK-------- 453
                :|      :| |             |::    :  |:           ||  |     :  | :  : :::|
Db        471 SKE-QLAQEAFV-----VSAMETTEKKLHGTANKLLSTVRETTRDVSGLHEKLDRKRAVE 524

Qy        454 -HNNNNNNNNDNNDNNN------------------DNNNNNNND-------------- 478
                ||:  : |          |                       ||: :|
Db        525 QHNSQVHENFAEQINRRFSVIQQTVDEYSVKQQGMLDFYTNSIDDLLGASSSALSATATA 584

Qy        479 ---SSSMVNNMINHMINNNINNNINVNNNNNNNNN------------------------ 510
                   | :|      ::    :::::    :            ::     :
Db        585 VAKSFASVQETVSKQVSHSVEEILKQETLSSQAKDDLQKLMTAHRTGLEQALRTDLLPVV 644

Qy        511 ----NNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRS 566
                    : |||  :|     |            |     |     |   :|||         : :|.   |||    |    :     |       |
Db        645 TAVLDLNSHLSHCLQSFLGVADKID-SHKEDMNSFFTEHSRS--LHKLRLDSSSAL---S 698

Qy        567 SCDNIMNKKKNNLHLAR--HSVGSKLTMFSYDPQKN-------KDNTFFKSNINKMEDNT 617
                 |  :        |   :  |:   || |      :  |       | |          :   :  |  |   |::::
Db        699 SIQSEYESLKEEIATAQSTHSEGVNNLISSLQNQLNLLAMETRQQFSGFLSKGGKLQESV 758

Qy        618 PKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINN 677
                     | :   :: :  :         :  :||   |:          :|          :     ||::::
Db        759 --GCLQQDLDLVSSDAIECI----SSHHSKFTEQSQAVTVEIRQ-----LAGSNMSTLEE 807

Qy        678 SN------MNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVN----EMDTSNKN 721
                 |:         ||||       |        :|       |: :  |:  |::       ::  |  :|:
Db        808 SSKQCEKLTNSINTICQESQQWCESAGQKMDSLLEEQVCYLHSSKKQIQTLHKD 861

RESULT 39
KF5C_HUMAN
ID   KF5C_HUMAN     STANDARD;      PRT;   957 AA.
AC   O60282; O95079;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin heavy chain isoform 5C (Kinesin heavy chain neuron-specific
DE   2).
GN   KIF5C OR NKHC2 OR KIAA0531.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [2]
RP   SEQUENCE OF 355-585 FROM N.A.
RA   Engelender S., Sharp A.H., Colomer V., Tokito M.K., Lanahan A.,
RA   Worley P., Holzbaur E.L.F., Ross C.A.;
```

```
RT   "Huntingtin associated protein 1 (HAP1) interacts with the p150Glued
RT   subunit of dynactin.";
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC       PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC   -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC       CHAINS.
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN, PROSTATE, AND
CC       TESTIS, AND MODERATE EXPRESSION IN KIDNEY, SMALL INTESTINE, AND
CC       OVARY.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC       TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC       KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC       ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC       DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC       INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC       VESICLES AND MEMBRANOUS ORGANELLES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB011103; BAA25457.1; -.
DR   EMBL; AF010146; AAD01436.1; -.
DR   HSSP; P56536; 2KIN.
DR   Genew; HGNC:6325; KIF5C.
DR   MIM; 604593; -.
DR   GO; GO:0005871; C:kinesin complex; TAS.
DR   GO; GO:0003777; F:microtubule motor activity; TAS.
DR   GO; GO:0006996; P:organelle organization and biogenesis; TAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    386       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      406    923       COILED COIL.
FT   DOMAIN      859    956       GLOBULAR.
FT   DOMAIN      174    315       MICROTUBULE-BINDING.
FT   NP_BIND      86     93       ATP (BY SIMILARITY).
FT   CONFLICT    355    360       TLKNVI -> STHASV (IN REF. 2).
FT   CONFLICT    583    585       EFT -> DRV (IN REF. 2).
SQ   SEQUENCE   957 AA;  109494 MW;  A9F25BB1C994322A CRC64;

  Query Match             6.4%;  Score 439;  DB 1;  Length 957;
  Best Local Similarity   21.4%;  Pred. No. 1.9e-12;
  Matches  216;  Conservative  173;  Mismatches  395;  Indels  226;  Gaps   41;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYI-ERHEFIVDKVFDD 63
                |||: | |||:| |   : |  |              |::| | |  |   :  :|:|
Db          9 IKVMCRFRPLNEAEILRGDKFI-------------PKFKGDETVVIGQGKPYVFDRVLPP 55

Qy         64 TVDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQ 121
                          ||    | :: |: |  || :   |||||| ||||:|| |          ||
Db         56 NTTQEQVYNACAKQIVKDVLEGYNGTI---FAYGQTSSGKTHTMEGK--LHDPQLMGIIP 110

Qy        122 YAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEV-VVKDLK 180
                | |||  :    |::     | :|::|||  |: |||    |    |:   |  |  ||
Db        111 RIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCT 170

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTS---LGKIAF 237
                 | : ||::   : :|    | : | : |: |||||:|  |::|  |   |      ||:
Db        171 ERFVSSPEEVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYL 230

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM-DSDKNHIPFRDSELTKVLR 296
                 :||||||: : | ::      :  |||:|| ||   | |  :  | |:|:|||::|::|:
Db        231 VDLAGSEKVSKTGAEG-AVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQ 289

Qy        297 DIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKN-------------KSTCINE 343
                |   |  :: ::    ||::     :| :||  :  |  |  ||              |   |
Db        290 DSLGGNCRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKE 349

Qy        344 EDDTNT-----------------------ERISILDSKGSEM--NASSIENVV------ 371
                ::  |                         |:||  |  |   |    |    |:|:
Db        350 KEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGI 409

Qy        372 -----------IKSNHLLSNNNNNKIN-RGKINDK-----IERNNILKN--KSFDKPREG 412
                           | |  :   :: :::|| : :: :|      ::::·:| :  : ::| :|
Db        410 STEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDYEKIQEE 469

Qy        413 FTS-TFGKYSSLNDIDKIKKNKKKGLINY--KSTLYNDNTINKKHNNNNNNNNDNNDNNN 469
                 |         :: ::: :: :   ::  :||  ||        | |          |     |
Db        470 LTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKT------RANEQLTDELAQKT 523

Qy        470 DNNNNNNDSSSMVNNMINHM------INNNINNNINVNNNNNNNNNNNSHNNHLPQPN 523
                       : |  :    : ||          ·| | :  ::
Db        524 TTLTTTQRELSQL-QELSNHQKKRATEILNLLLKDLGEIG------------------- 562

Qy        524 YAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIK-LKNRS----SCDNIMNKKKNN 578
                       |:|  :| |:|   :               |:    ::  :| | |||        |:| |
Db        563 -GIIGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKSLVNRSKQLESAQMDSNRKMN- 620

Qy        579 LHLARHSVGSKLTMFSYDPQKNKDNTFFKSNI----NKMEDNTPKDILYESRNVSNMNGN 634
                  :   :   :  |       | |   |  :   |:        ::|::  :| |   |    :: :
Db        621 --ASERELAACQLLISQHEAKIKSLTDYMQNMEQKRRQLEES--QDSL--SEELAKLRA- 673

Qy        635 VLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIY 694
                      :     |::| :|       |     :    :   :     :      | |     ::    |
Db        674 ------QEKMHEVSFQD------KEKEHLTRLQDAEEMKKALEQQMESHREAHQKQLSRL 721

Qy        695 KSNYNSNQSISDVQIRYVN---EMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNN 751
                :         | | | :|| :|     :::       ::|
Db        722 RDEIEEKQKIID-EIRDLNQKLQLEQEKLSSD--------------------------- 752

Qy        752 NNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNY 811
                      |·: :|:                     :  |            |:|       :      :    |:
Db        753 ---YNKLKIEDQEREMKLEKLLLL------------NDKR----EQAREDLKGLEETVSR 793

Qy        812 SMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNN--NIVI 869
                :     | ||   |  :: ||   : :|          :       :|:|  :|  ||   :
Db        794 ELQTLH-NLR-KLFVQDLTTRVKKSVELDNDDGG-----GSAAQKQKISFLENNLEQLTK 846

Qy        870 VNNNMGNNNS------PRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKD 913
                |:  :  :|:         |:::  |        : : :|   |       | |:|  |:|
Db        847 VHKQLVRDNADLRCELPKLEKRL---RATAERVKALESALKEAKENAMRD 893

RESULT 40
KINH_DROME
ID   KINH_DROME     STANDARD;      PRT;   975 AA.
AC   P17210; Q9V7L9;
DT   01-AUG-1990 (Rel. 15, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin heavy chain.
GN   KHC OR KIN OR CG7765.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89168428; PubMed=2522352;
RA   Yang J.T., Laymon R.A., Goldstein L.S.B.;
RT   "A three-domain structure of kinesin heavy chain revealed by DNA
RT   sequence and microtubule binding analyses.";
RL   Cell 56:879-889(1989).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   MUTAGENESIS.
RX   MEDLINE=93030741; PubMed=1384131;
RA   Gho M., McDonald K., Ganetzky B., Saxton W.M.;
RT   "Effects of kinesin mutations on neuronal functions.";
RL   Science 258:313-316(1992).
CC   -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC       PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC   -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC       CHAINS.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC       TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF
CC       KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC       ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC       DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC       INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC       VESICLES AND MEMBRANOUS ORGANELLES.
CC   -!- MISCELLANEOUS: MUTANT FLIES DISPLAY IMPAIRED ACTION POTENTIAL
CC       PROPAGATION AND NEUROTRANSMITTER RELEASE AT NEUROMUSCULAR
CC       JUNCTIONS, BUT ARE STILL CAPABLE OF TRANSPORTING CERTAIN
CC       MEMBRANES, INCLUDING SYNAPTIC VESICLES, TO THE NERVE TERMINAL.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC       SUBFAMILY.
CC
CC   --------------------------------------------------------------------------
```



```
DR   EMBL; M24441; AAA28652.1; -.
DR   EMBL; AE003807; AAF58029.1; -.
DR   PIR; A31497; A31497.
DR   HSSP; P33176; 1BG2.
DR   FlyBase; FBgn0001308; Khc.
DR   GO; GO:0005871; C:kinesin complex; IDA.
DR   GO; GO:0008017; F:microtubule binding activity; IDA.
DR   GO; GO:0003774; F:motor activity; NAS.
DR   GO; GO:0007317; P:regulation of pole plasm oskar mRNA localiz. . .; IMP.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    334       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      335    931       COILED COIL.
FT   DOMAIN      932    975       GLOBULAR.
FT   DOMAIN      180    321       MICROTUBULE-BINDING.
FT   NP_BIND      92     99       ATP (BY SIMILARITY).
FT   CONFLICT    515    515       A -> T (IN REF. 1).
SQ   SEQUENCE   975 AA;  110399 MW;  24840EF414790888 CRC64;

  Query Match             6.4%;  Score 439;  DB 1;  Length 975;
  Best Local Similarity   21.9%;  Pred. No. 1.9e-12;
  Matches  219;  Conservative  164;  Mismatches  429;  Indels  186;  Gaps   37;

Qy     5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 64
         |||| | |||:: |:|     ::   ||    ::|    :   |       : ||||
Db    13 IKVVCRFRPLNDSEEKAGSKFVVKFPNN----VEENCISIAGKVY------LFDKVFKPN 62

Qy    65 VDNFTVYENTIKPLIIDLYE--NGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY 122
              ||    | :: |:    || :   |||||| ||||:|| |    | |   ||
Db    63 ASQEKVYNEAAKSIVTDVLAGYNGTI---FAYGQTSSGKTHTMEG--VIGDSVKQGIIPR 117

Qy   123 AAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEV-VVKDLKI 181
            |||  :   : :    | :|:||||  |: |||   |:  ::   |  |  ||
Db   118 IVNDIFNHIYAMEVNLEFHIKVSYYEIYMDKIRDLLDVSKVNLSVHEDKNRVPYVKGATE 177

Qy   182 LRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDIN----KNTSLGKIAF 237
           | : |::   : :|   | | | : |: |||||::  |::|  |    |  | ||:
Db   178 RFVSSPEDVFEVIEEGKSNRHIAVTNMNEHSSRSHSVFLINVKQENLENQKKLS-GKLYL 236

Qy   238 IDLAGSERGADTVSQNKQTQTDGA-NINRSLLALKECIRAM-DSDKNHIPFRDSELTKVL 295
         :||||||: : | ::   |  | | |||:|| ||   | |: | :| |||:|||:||::|
Db   237 VDLAGSEKVSKTGAEG--TVLDEAKNINKSLSALGNVISALADGNKTHIPYRDSKLTRIL 294

Qy   296 RDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISIL 355
         ::   | ::: ::   ||      :| :|| :  | |  ||   |:|||     |
Db   295 QESLGGNARTTIVICCSPASFNESETKSTLDFGRRAKTVKN-VVCVNEE--LTAEEWKRR 351

Qy   356 DSKGSEMNA---SSIENVVIKSNHLLSNNNNNKINRGKINDKIERN-------------- 398
           |  | ||     :| : |:     :        :  : | :| :
Db   352 YEKEKEKNARLKGKVEKLEIELARWRAGETVKAEEQINMEDLMEASTPNLEVEAAQTAAA 411

Qy   399 ------------NILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYN 446
                     |:  : : ::     |     |  |:| |: : |::     |   |
Db   412 EAALAAQRTALANMSASVAVNEQARLATECERLYQQLDDKDE-EINQQS---QYAEQLKE 467

Qy   447 DNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNN 506
              ::   |     :   :        |  :   |  ::  :    :| :   :   :
Db   468 QVMEQEELIANARREYETLQSEMARIQQENESAKEEVKEVLQALEELAVNYD-QKSQEID 526

Qy   507 NNNNNNNSHNNHLPQPNYAFTDTS-DFSSLDDMNCHLNNNDKSIFLHKKNLRD------- 558
         | | : :: |  | |    |   | :   | ||: |       :  :   |||
Db   527 NKNKDIDALNEELQQKQSVFNAASTELQQLKDMSSHQKKRITEMLTNL--LRDLGEVGQA 584

Qy   559 ------NIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINK 612
               :| || :      :| : :  :|| :    |      |:  :    ::: ||
Db   585 IAPGESSIDLKMSALAGTDASKVEEDFTMARLFISKMKTEAKNIAQRCSNMETQQADSNK 644

Qy   613 MEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKD------ENHNDNKINNGVINI 666
```

```
                    ||:             |  | :::   |:  |       |  |  :     |:
Db      645 KISEYEKDL-----------GEYRLLISQ---HEARMKSLQESMREAENKKRTLEEQIDS 690

Qy      667 INNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSN-QSISDVQIRYVNEMDTSNKNNDNI 725
            :           :::::|           :| ::|    : :   | |:|:
Db      691 LREECAKLKAAEHVSAVNAEEKQRAEELRSMFDSQMDELREAHTRQVSEL---------- 740

Qy      726 FFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQ 785
                                            : |  : :   :   :   ||    :|
Db      741 ----------------------------RDEIAAKQHEMDEMKDVHQKLLL--AHQQMT 769

Qy      786 PDNNK-------NTSNIQN-INTNKNNQ---------DGNVNYSMNFCHYNLNDKNYLID 828
             |  |          :| :|| | ||: :         :  |   :   | ||  | :: |
Db      770 ADYEKVRQEDAEKSSELQNIILTNERREQARKDLKGLEDTVAKELQTLH-NLR-KLFVQD 827

Qy      829 LNNKEQKDKNIHGCDNNIIQNRND-----FEKKKKTNFYNNN--NIVIVNNNMGNNNS-- 879
            |  : :|           |::   ::       :|:| :|  ||   :  |:  :  :|:
Db      828 LQQRIRK---------NVVNEESEEDGGSLAQKQKISFLENNLDQLTKVHKQLVRDNADL 878

Qy      880 ----PRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKD 913
                |:::  |      ::: :|  |    | |:  |:|
Db      879 RCELPKLEKRL---RCTMERVKALETALKEAKEGAMRD 913
```

Search completed: October 2, 2003, 16:17:44
Job time : 44 secs

```
OM protein - protein search, using sw model

Run on:         October  2, 2003, 16:14:11 ; Search time 52 Seconds
                                             (without alignments)
                                             2382.024 Million cell updates/sec

Title:          US-10-006-780-2
Perfect score:  6842
Sequence:       1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_76:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	789.5	11.5	4550	2	T18440	hypothetical prote
2	778	11.4	1844	2	D71612	hypothetical prote
3	765	11.2	1619	2	T18499	hypothetical prote
4	755	11.0	668	2	T02791	mitotic centromere
5	752	11.0	3394	2	T18501	hypothetical prote
6	748.5	10.9	1817	2	D71606	hypothetical prote
7	732	10.7	1254	2	T18277	kinesin heavy chai
8	728	10.6	716	1	A44259	kinesin-related pr
9	722	10.6	671	2	T10755	kinesin-related pr
10	698	10.2	1436	2	D71618	hypothetical prote
11	696	10.2	807	2	B71605	hypothetical prote
12	676	9.9	2150	2	S71629	sensory transducti
13	656.5	9.6	747	2	T23607	hypothetical prote
14	646	9.4	1256	2	S14556	asparagine-rich pr
15	611	8.9	1711	2	T18429	hypothetical prote
16	603.5	8.8	2523	2	T18477	hypothetical prote
17	601.5	8.8	4981	2	T18489	hypothetical prote
18	597	8.7	537	2	A23770	asparagine-rich pr
19	591	8.6	3724	2	T18427	hypothetical prote
20	587	8.6	2485	1	H71621	serine/threonine-s
21	584	8.5	1650	2	T18444	hypothetical prote
22	581	8.5	2708	2	T09079	probable chloroqui
23	581	8.5	3844	2	T18402	asparagine/asparta
24	579	8.5	2380	2	E71604	hypothetical prote
25	577.5	8.4	2500	2	G71609	hypothetical prote
26	569.5	8.3	2010	2	B71616	phosphatase (acid
27	567	8.3	2819	2	T09080	probable chloroqui
28	563.5	8.2	3973	2	B71612	hypothetical prote
29	561.5	8.2	1111	1	A42640	kinesin-related pr
30	560.5	8.2	2339	2	A45597	DNA-directed RNA p
31	552	8.1	2539	2	B71619	hypothetical prote
32	549	8.0	2510	2	T28160	hypothetical prote
33	544	8.0	2269	2	T18472	hypothetical prote
34	543	7.9	1984	2	A44396	P-type cation tran
35	542.5	7.9	1553	2	T18502	hypothetical prote
36	541.5	7.9	1073	2	S14032	kinesin-related pr
37	538.5	7.9	1085	2	T38378	kinesin-like prote
38	537.5	7.9	1714	2	E71609	Ser/Thr protein ki
39	534.5	7.8	1516	2	E71619	RAD2 endonuclease
40	534	7.8	1058	2	T47525	kinesin-related pr
41	533.5	7.8	2391	2	T18410	carbamoyl-phosphat
42	533	7.8	1390	2	T14004	trfA protein - sli
43	532	7.8	813	2	T46242	kinesin-like prote
44	529.5	7.7	669	2	S14535	asparagine-rich pr
45	528	7.7	1166	2	H71609	hypothetical prote

ALIGNMENTS

```
RESULT 1
T18440
hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18440
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18440
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4550 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB11121.1
C;Genetics:
A;Map position: 3
A;Note: C0425w

  Query Match             11.5%;  Score 789.5;  DB 2;  Length 4550;
  Best Local Similarity   21.0%;  Pred. No. 2.7e-25;
  Matches  380;  Conservative  247;  Mismatches  498;  Indels  683;  Gaps   70;

Qy         11 KRPLSELEKKKKDSDIITVKNNCTLYID-EPRYKVDMTKYIERHEFIVDKVFDDTVD-NF 68
              |:  :: :    |||  |           |  | :|    | ||  :       | ::||  | ||
Db        795 KKDDNKKDGDKKDDD----KKGDDKYDDGDDKYDDDDNKYDDD-----DDIYDDDDDFNF 845

Qy         69 TVYENTIKPL-----IIDLYENGCV--------CSCFAYGQ--TGSGKTYTM-------L 106
                 | | |       ::    : | :          ||     |     | | | |          :
Db        846 DHDSETSKRLSNYDSLLKDKKKGSILKKEDTLKCSKMINVQEKEKSSKMYCMNTNNCNVV 905

Qy        107 GSQPYGQSD-------------------TPGIFQYAAGDIFTFLNIYDKDNTKGIFISFY 147
                :  |  :|                      :||  :       |:  |::   |            |
Db        906 KEKNYFVNDYMHIINNLSDHEMNINMMKSPG-YNNLFKDVEHFIHKRD----------FI 954

Qy        148 EIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRVLTKEEL-----------ILKMI 195
                 :      ||       |  |::   |:|   |:|   |        :  ||              :
Db        955 NLNILKLI----KEKIIQVLQNNYKQVCVFCKKTTGIKTKCMFPSCSTYFHIYCYYNKYM 1010

Qy        196 DGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQ 255
                |   :|   :    :   : ||         ||        :             |||:     :
Db       1011 QHVRKKKNDLQCASQKKSR-------------KNQRAASV-------KRRGAE-----GK 1045

Qy        256 TQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTI 315
                : ||:                   |::|  |: :   |  ::   :     |       |:
Db       1046 DKKDGS-------------RSLDRKKDKVTRRSNKYEESANDSDGKKNNG---------- 1082

Qy        316 SCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSN 375
                       :   :    :        |  :   |::|   |     |    :    |     |     |:    |
Db       1083 -------DDKKKGDNIGKNNNNNNNNNDDDSNNNGDDSNNNGDDSNNNGDGNNNINGDGN 1135

Qy        376 HLLSNNNNNKIN-------------------------RGKINDKIERNNILKNKSFDKP 409
               : :: :  || ||                          ||: |:     : :|  ||   |
```

```
Db     1136 NNINGDGNNNINGDGNNNNDCLEEKNQGDNQKKRKRNKRGRNNNDDTSSLVLLNK--DDS 1193

Qy      410 REGFTSTFGKYSSLNDID-----------KIKKNKKKGLINYKSTLYNDNTINKKHNNN 457
            |:   : :||  |  : |           |:||:|: |  |         ||: ||   |::
Db     1194 RQKKKNKYGKDHSDQEKDSPSKINNRSKRNKVKKDKEVGNI-------NDDKINNNTNDD 1246

Qy      458 NNNNNDNNDNNNDN---NNNNNNDSSSMVNNMIN-HMINNNIN---------------- 496
            | ||| | || |:|   :| |||  :   :|| ||  |||||
Db     1247 NKNNNINGDNKNNNINGDNKNNNINGDKINNNINGDKINNNINGDKISKNYHSHNYDNEE 1306

Qy      497 ------NNINVNNN------------NNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDD 537
                  || | |              |:  || ||   :   | :  |:          || |
Db     1307 ITHEQKNNFNKNKNTKRYRKSINDKEMNDILNNKNNDTDKKLNKYNFLMEYQKIISS-DK 1365

Qy      538 MNCHLNNNDKSIFLHKKNLRD--------NIK------------LKNRSSCDNIMNKKK 576
            :   ::|| |  |     ||::|         |||         :|:  :   || : |
Db     1366 ITSGISNNMKDI----KNIKDIKDIKDIKNIKDIKDIKDIKDIKDIKDIKNVKNIKDIKS 1421

Qy      577 N-----------NLHLARHSVGSKLTMFSYDP--QKNKDNTFFKSNINKMEDNTPKDI 621
            |            ::|| :  | :|    :  |   || :|    |: ||  |        :
Db     1422 NEDIKDIKDVNDVSSVHLNK-CVNNKSDTQDFCPSDDKNINNIIKKNKINIKTDQVYNPL 1480

Qy      622 LYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMN 681
            :  | : :: | |                 | :        ||    : ||  ::| |: ||:| |
Db     1481 VDSSSSSTDCNKNY-------------KKKEKAVKKIGWNIKNIKLHNNNNNNNNNNN 1526

Qy      682 SINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFD-----------A 729
            : ||:|::||:: : :|      |  :      ||  ::| | |:|::|             :
Db     1527 NNNNNNISSNNVVQEDYKMKGKGSHKK----NEKHSTNYNQDDIYYDNDFSSICSLLSSS 1582

Qy      730 ISCDNNM---------------------------------------------------- 736
            :| :|::
Db     1583 LSTNNDIDSSYVSNNSLYNSSSMFSNDDCFSLCYSSVSSSYEYDDINNVNLKNKIHNFPS 1642

Qy      737 -YPNITNNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSH---NLFQPDNNKNT 792
             | |  ||||||||||||||| :  |||  :     :|  |    :      :|:   :| :
Db     1643 FYKNELNNNNNNNNNNNNNNNNNNECHNYNRKHHCTHSRDLSCSSRSFWGSLYNTSDNISE 1702

Qy      793 SNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLI-----------------------D 828
            | |  |     : | : ||  :   : : :|:  |                        |
Db     1703 SEIS--NNLSFSSDKSENYCSSISDHEI-EKDVTIYRNTYHRIINRWKDGKKNTGSKKRD 1759

Qy      829 LNNKEQ----KDKNIHGCDN--NIIQNR------------------NDFEKKKKTNFYN 863
            ::||:        |||  |||  ||  |:                  :  ||::  |   |
Db     1760 MSNKKGDKSCNKKNI--CDNKKNIYDNKKLIYDNKDSSTDAPVENIDHLIEKEEIDNILN 1817

Qy      864 NNNIVIVNNNMGNNNSPRM---KYGLCGSHTS--------------------------- 892
            |   :     |||:   |  ||:|::
Db     1818 NLGLYTYNLKNNNNNNVYMKTIKLTLCCTHSNEHDTKAVMNIKLNTLNNKKYNNINCIFY 1877

Qy      893 -----------------------------------------------IDNMKNNEMKNNEMK 907
                                                              ||| ||:  ||:
Db     1878 HYYNSFLHSLFYSDEYIELIRKEERKEDIEKEKIKRGKYCHQEDRDDNMNND--NNDNN 1935

Qy      908 DNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQND-------------YCHN--DN 952
            :|   :|:  :|||| ::::  :|||  ||  :|    |                :  ||   |:
Db     1936 NNNDNNNNDNNNNNDNNNNDNNNNDNNNSGEDQLINEDNKKRKINNLSTHEHTHNMSDH 1995

Qy      953 TFTIRRKNN--------------------------TNINSNIYQNDDIIYTINSLNDY 984
              ::|  |                            || :: |  :  |    ::  |:
Db     1996 NKEPKKKKNKSNETQINDDTNKKETHKNEIDHKEQKDTNCDNGIEEKKD---QVSKSNNI 2052

Qy      985 MSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDD-----QDKYDDNDNNNVDNNNKN 1039
            :|        |:         :  :|  ||   |:|  |:  ::  |||  |||  ||| |
Db     2053 ISKVFHFFERKNN----NIKKDEHKKEYNERNDKMSECDMLLQNKRDNDHNNN--NNNNN 2106

Qy     1040 NVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNN-------EYLSY-- 1090
             : |||:|||:|:|:|||:|:   |||:|:|||  :|  : :||       |  ||
Db     2107 DDDNNNNNNNDDDNNNNNNNNDDHNNNNNNNDDDNNNNNNNDDDNNKDPTLDGETRSYID 2166

Qy     1091 ---------FQKNVDTIINN----CL---------------NSLDISSMYDDTKEILNN 1121
                     | | :| || |      |:                |  |:|          ::: ||
Db     2167 EINQKEINIFSKYIDKIIENVETKCIMKNNVCDGIYYLIRQENGLNI------CQKVNNN 2220

Qy     1122 ILLSKYKAEK-----------DNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKL 1169
            |: :| |  ||              |    |   |||       :     :|       :|| : | :
Db     2221 IIKNKMKQEKFYKKKKQNKNNHKNKNKNKYNYDINIFNNVPVD------IQKRNYINTNI 2274

Qy     1170 LLLFKKNVDTQINNET-----------SDLRKDLVMCHICNNNPDDQFHFYAYSRLE--- 1215
             : :    |  :   :| :              |||:| :         | |         : | : ||
Db     2275 ININNNNNNSNSNHSNHNNNSNHINLSDLKKKI------NINK----YTYLFYRLPLFL 2324

Qy     1216 ---------KDIINLIMLRQIWCESENL-----RLLYQFLVVEYQNKSANSVLLN----- 1256
                     |: :: |::  :  :  ::        :||   :::   | : | :  |
Db     2325 FGDKYINDIDDVKSIYMIKNFFNQYNDVTNTSNNILYVIDLLKKYNTNHNILKKNNMIQS 2384

Qy     1257 ----------------------------------VSSNNGDI---ILLNKKLVQDNIKNS 1279
                                               |  |  ::     :  :    :|::| |:
Db     2385 PKNYICEKEEQSNDPAKNTHVPFNHNALDTSHPPVDTTNTNLLSDVNIETNEIQEHINNN 2444

Qy     1280 M--DHNNI 1285
            :  ||||:
Db     2445 IRNDHNNV 2452
```

RESULT 2
D71612
hypothetical protein PFB0540w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: D71612
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: D71612
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1844 <GAR>
A;Cross-references: GB:AE001401; GB:AE001362; NID:g3845209; PIDN:AAC71897.1; PID:g38452
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0540w

```
  Query Match             11.4%;  Score 778;  DB 2;  Length 1844;
  Best Local Similarity   24.5%;  Pred. No. 3.2e-25;
  Matches  340;  Conservative  202;  Mismatches  439;  Indels  406;  Gaps   66;

Qy      205 VNSQNDESSRSHAILNI-DLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQT----- 258
            || :|| :         :  : :|| ||:  : :  |:|:            :| ::  |   |
Db       37 VNRRNDMNYEMYNSINSGHMSNINNNTNNLQDAYINKELHYMNSDKINIAKNHQNVHMTA 96

Qy      259 -----DGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISP 313
                 |  | | :::  |    |   |  |:|:   |: |         :      : ::   |:
Db       97 TYNNMDKNNANNNIIHLNNHIN-MTDDQNY--FYNSTTNNKM-------NNTLKENNFNN 146

Qy      314 TISCCEQTLNTLRYSSRVKNFKNKSTCINEEDD-TNTERISILDSKGSEMNASSIEN--- 369
            ::      |:|   | :   || | :   ||:|    |   |:  :::  :  |  :: |
Db      147 NMN----TVNNSFYHNTDNNFLNFTRNQNEQDTYVNNNIINNFNNQNVDKNINNNNNNPN 202

Qy      370 ----VVIKSNHLLSNNNNN----KINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYS 421
                :  |   ||:    :   |      | ||   :|    |    :           |
Db      203 KNVESINKFNHIYNMQNFNHFIPNISNGK-NGNLENNASLSH-----------NVNSVS 249

Qy      422 SLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNDNDNNNDNNNN---NNND 478
            ::::     |        |   :   |  |:  | ||  | |  |  | |||  ||  |||
Db      250 TISELHNFNYMNNMDLNNVDMNNMNMNNMNMNNMNMNNMNMNNMDMNNVNMNNMNMNNMD 309

Qy      479 SSSM-------VNNM-INHMINNNIN-NNINVNNNNNNNNN-----------NNNNSHN-N 517
             :||       :|||  :|:|   ||:| ||::|||  :|| ||                 | |||:  |
Db      310 VNSMNMNNMDNMNNMNMNNMDVNNMNMNNMDVNNMDNMNNMDNMNNMKNLSNFNNSYQYN 369

Qy      518 HLPQPN----------------YAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKN------ 555
            :||  |                      :   : |:||:     | ||:|::    |:  |
Db      370 SIPQFNSSSRFNNITHFNNGISHNVNNVSNFSN----NAHLDNSNMNRLNAVNNFGDIN 425
```

```
Qy    556 -----------LRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQ------ 598
Db    426 SFHDPLNEMQVLNKNVNMQNENHF-NVMNDEMKNYNNVK-----RINSISHIPYMNNLKN 479

Qy    599 --------KNKDNTFFKSNINKMEDNTPKDI---LYESRNVSNMNGNVLLG-LNKNTHHD 646
Db    480 YNEHTSMVKGKGNTNRKKSNNLKINNNPGSVNARAISENNQSTAHGNIPMGSVDKVIKHD 539

Qy    647 ISTKD-----ENHNDNKIN-----NGVINIINNSNVNSINN-------SNMNSINN-SNM 688
Db    540 RMDNDLKNINNMNMNMSMNNMNSMNNMNMINMNNMNMNMNMINMNNMNMNMNMNMSNM 599

Qy    689 NSNSIYKSNYNSNQSI---------SDVQIRYVNEMDTSNK---------------NND 723
Db    600 NNTSILN---NNNKKITKRGRAKKNSTININNINKMNSTNNKSSMINMNSVNNMNSVNNM 656

Qy    724 NIFFDAISCDN-NMYPNITNNNNNNNNNNNNNNNNI-DVENYNNRDGTNNSMKLYAYNS- 780
Db    657 NSVKNMNSVNNMNSVNNINNVNNINNVNNINNVNNINNVNNINNVNNINNVNNINNVNNM 716

Qy    781 HNLF-----------QPDNNKNTSNIQNIN---------TNKNNQDGNVNYSMNFCHYNL 820
Db    717 NNMFNVNPQLNIMGIMKDINNNNITVSNKNKLMNNYINDNNIMNMEGSINETYNF-DGTL 775

Qy    821 NDKNYLIDLNNKEQKDKNIHGCDN------------------NIIQNRNDFEKKKKTNFY 862
Db    776 NNKN-VSNNNNNDIHDKGVLNTLNRSKSSSYIKPHRNLTLPINMYMNNTYMYNSKAYVNY 834

Qy    863 NNNNIVIV-----NNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIK 917
Db    835 ENQNWMAQQDCNDKNGLSLNEGPRYN----------DNNNNNNNNNNNNNNNNNNNN-N 883

Qy    918 SNNNNSSSSSSSNNNIYNNINDDDTFQNDY----CHNDNTFTIRRKNNTNIN-------S 966
Db    884 NNNNNNNNHSIINNNITQGIHRGSMINNQHFDELSYNPNGIFLEKNTITNYNEMLGSGYN 943

Qy    967 NIY---------------QNDDIIYTINSLNDYMSN---------------------TLL 990
Db    944 NMYDKNAVKGNMNLIGKHSNYDLMKNGNMINGYKLNMQEVQKGMNEVGKKRAPRRRAKTI 1003

Qy    991 HFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDN----------------NNVD 1034
Db   1004 NLKSINRFNTVSLG-------LKGNH-NMENMNNMDNMDNMNNMINVSNVNNVNNMNNVN 1055

Qy   1035 N-NNKNNVD--NNVDN-NNVD--NNVDN-NDKNNVDN----NNVDNDDDDVDFHNIKNFN 1083
Db   1056 NVNNMNNVNNMNNVNNMNNVNSVNNMNNTNNMNNMNNVNSVNNMNNTNNMNNMNNTNNMN 1115

Qy   1084 NNEYLSYFQKN-------VDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIK 1136
Db   1116 NMNRMNHIDNNMVTNLNYMDNKINNAGNNLN----GEFSKRILYN--------------- 1156

Qy   1137 KYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSD-----LRKD 1191
Db   1157 -------RSKSTENIKEMMQPHNEANNI----------------SGNNTSDSNNITLNKN 1193

Qy   1192 LVMCHICNNNP--DDQFH---FYAYSRLEKDIINLIMLRQIWCE--------SENLRLLY 1238
Db   1194 LVEMILRNNKPSIDKNIHESVNRSYTSFLMNIGSSYLKKKKTAEIKTGENNTENNKGIVN 1253

Qy   1239 QFLVVEYQNKSANSVLLNV------------SSNNGDIILLNKKLVQDNIKN-----SMD 1281
Db   1254 VNSQVEEKGKSENKSILEIKKEGQIKNVIYNNNNNNNEKDQNADQYQDQNKNHKQDQRQD 1313

Qy   1282 HNNIHKK 1288
Db   1314 QNKSHKQ 1320
```

RESULT 3
T18499
hypothetical protein C0770c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18499
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18499
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1619 <LAW>
A;Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331920; PIDN:CAB11138.1
C;Genetics:
A;Map position: 3
A;Introns: 87/3
A;Note: C0770c

```
  Query Match             11.2%;  Score 765;  DB 2;  Length 1619;
  Best Local Similarity   22.0%;  Pred. No. 9.6e-25;
  Matches  369;  Conservative  246;  Mismatches  568;  Indels  496;  Gaps   66;

Qy      5 IKVVVRKRPLSELEKKK-KDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db     16 IKVIVRCRPLNEKEKNDINNEEVVRINNN------EVILTINRNNEIYEKKYSFDYACDK 69

Qy     64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-----QPYG------ 112
Db     70 DVDQKTLFNNYIYQ-IVDEVLQGFNCTLFCYGQTGTGKTYTMEGKILEHLKQYDNNKKVD 128

Qy    113 ---------------QSDTPGIFQYAAGDIFTFL---------------NIYD-KDN-- 138
Db    129 LNESINSDISYCYELCENEDTGLIFRVTKRIFDILNKRKEEKIRHFKGKKNIFDEKDNEH 188

Qy    139 -------TKG-------------------------------------------------- 141
Db    189 INVINKETNGNNNNIINNNYNNLLCNKINIDKNDIYNNLKINKEPNNLLVNEKSRDITCD 248

Qy    142 ----------------------------IFISFYEIYCGKLYDLL------QKRKMVAAL 167
Db    249 VVQKCSVDKNRNNDMYQTFDDKNMYDFNIKISYLEIYNEELCDLLSSTNENMKLRIYEDS 308

Qy    168 ENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDL--KD 225
Db    309 NNKSKGLNVDKLEEKSINSFEEIYYIICSAIKKRRTAETAYNKKSSRSHSIFTITLIIKD 368

Qy    226 IN----KNTSLGKIAFIDLAGSERG-ADTVSQNKQTQTDGANINRSLLALKECIRAMDSD 280
Db    369 INNVGESITKIGKLNLVDLAGSENALKSSYGSLKIRQQESCNINQSLLTLGRVINSLIEN 428

Qy    281 KNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTC 340
Db    429 SSYIPYRDSKLTRLLQDSLGGKTKTFIVATISPSSLCIDETLSTLDYVFRAKNIKNRP-- 486

Qy    341 INEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNI 400
Db    487 --EINIKTTKQLKIKD----------LNNEIEKLRNAL---NLSREKRGVYLDTEEYNNI 531

Qy    401 ----LKNKS---------FDKPREGFTSTFGKYSSLNDI---------DKIKKNKK---- 434
Db    532 QNSLKKNKEILLEKEKILFEKSKK-IKNLLNKMDNTDDLQNKIFFFLKDTIQKYKNIQAF 590

Qy    435 -----KGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVN----- 484
Db    591 HQILIKKIIEEK--YITQFLLNKYHSLEENFINQKKSFQQIKTNMSNIMHDNLLNIKKNI 648

Qy    485 NMINHMINNNINNNINVNNNNNN----NNNNNNSHNNHLPQ----PNYAFTDTSDFSSLD 536
Db    649 HQENHLLNDICKTILFILDDTKNYLVQNKKSYNDYINHFEELNSNINFKYNDTLNF-ILD 707

Qy    537 DMNCHLNNNDKSIFLHKKNLR--------------DNIKLK------------------- 563
Db    708 HMDI-LQKYDHDYFSLSENIKKEISQCKINIPHISDEAKQKMPAIIGVYNTNGLKKEKQS 766

Qy    564 ----------------NRSSCDNIMNKKKNNLH-LARHSVGSKLTMFSYD---------- 596
```

```
Db        767 GEKIICEKQSGEKIICEKQSGEKIICEKQNGEEIICEKQSGEKIICEKQNGDEIICEKQN 826

Qy        597 ------PQKNKDNTFFKSNINKMEDN-TPKDILYESRNV-SNMNGNVLLGLNKNTHHDIS 648

Db        827 GEGIICEKQNGEGKTCQKQYGKEENNEMPFDHIEESFNYPENKEFIKDLLLNRQGEDIIQ 886

Qy        649 TKDENHNDNKINN---------GVINI---------------------------INNSN 671

Db        887 NLDNHESEKKLYNAKSLIQYIKSICNIFHIFLNESFMLLEKKCHEKYIHWDKDKKIMNDN 946

Qy        672 V--------------NSINNSNMNSINN--SNMNSNSIYKSNYNSNQSISDVQIRYVNEM 715

Db        947 IEYIYKKYDVFEKEDNIIINNYKKMINEDIQEFKENILKDIELLINKNVEDIYEKMNNKF 1006

Qy        716 --------DTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNNRD 767

Db       1007 DLLNKELYDKSKKNYKSVFMNSVKKIDDTIQHFSECNINDEKEYNNINNDI------KRK 1060

Qy        768 GTNNSMKLYAYNSH-----NLFQPDNNK---NTSNIQN--INTNKNNQDGNVNYSMNFCH 817

Db       1061 ITNYHTVLYNILSYLKRKSYIEEKEHSKCLNDTLNIYNNLLYTNNLNIDKIIN------- 1113

Qy        818 YNLNDKNYLIDLN-NKEQKDKNIHGCDNNIIQNRNDFEKKKKTN-----FYNNNNIVIVN 871

Db       1114 ------NFTSQTNINKIEKRKLYNVISEKILQNKNKESQEQKGGRRIHMLKSNENIKLFN 1167

Qy        872 NNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSN- 930

Db       1168 NIFE------------------QNLNKCNMYIAELIDILKKKDYLQDIPSTSSYDLMENM 1209

Qy        931 -NNIYNNINDD-----DTFQN----------DYCHNDNTFTIRRKNNTNINSNIYQNDDI 974

Db       1210 DDEFYLNIKDDVQRIRETFVNVIDMFTPMKITSLNRDSTY------DTNYDNN--KNDD- 1260

Qy        975 IYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDN------ 1028

Db       1261 ----NNKNNHDNNKNNHDNNKNNH---DNNKNGDNNKNDHDNNKNGDNNNYDDNKMDQGH 1313

Qy       1029 DNNNVDNNNKNNVDNNVDNNNVDNNVDNN-DKNNVDNNNVDNDD---DDVDFHNIKNFNN 1084

Db       1314 DHNNYLYNDDNKMDQGHDHNNYLYNDDNKMDQGHDHNNYLYNDDNKMDQGDHHNYYLYND 1373

Qy       1085 N------EYLSYFQKNVDTI-------INNCL----------NSLDISSMYD-------- 1113

Db       1374 DNKIDQGDHHNYYLYNDDIIRQNLYLEVKNTLMKEIEQITTTHDIDLSKVFEQIDENFQF 1433

Qy       1114 ---DTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLL 1170

Db       1434 LIQDDKIVAHENNYSRYNSLNDHTTNSSIPHNINNKG----DKNKLNVKTNHNNLPLNEY 1489

Qy       1171 LLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINL-IMLRQIWC 1229

Db       1490 MLYNKNT---TDNPIQPIDETYQELNYLNNNQHNK-------RFDVNTENMSVTSLKEYE 1539

Qy       1230 ESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKLVQDNIKNSMDHNNIHKK 1288

Db       1540 RKRNINVQNVHITNNYSNNNNNNNNNNNNNNNNNNN----NNKFQSTKIK-AIENRAFPKK 1593
```

RESULT 4
T02791
mitotic centromere-assocated kinesin MCAK [imported] - Leishmania major (strain Friedlin
C;Species: Leishmania major
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 19-May-2000
C;Accession: C81455; T02791
R;Myler, P.J.; Audleman, L.; deVos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.;
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999
A;Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c
A;Reference number: A81455; MUID:99178987; PMID:10077609
A;Accession: C81455
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-668 <PYL>
A;Cross-references: GB:AE001274; NID:g3264850; PIDN:AAC24615.1; PID:g2978452; GSPDB:GN0
A;Experimental source: strain MHOM/IL/81/Friedlin
C;Genetics:
A;Gene: MCAK
A;Map position: 1
A;Note: MCAK
C;Superfamily: kinesin-related protein KIF2; kinesin motor domain homology
F;135-459/Domain: kinesin motor domain homology <KMOT>

```
  Query Match             11.0%;  Score 755;  DB 2;  Length 668;
  Best Local Similarity   42.9%;  Pred. No. 9.8e-25;
  Matches  172;  Conservative   67;  Mismatches  134;  Indels   28;  Gaps    8;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFD 62

Db        133 SRILVAIRKRPLSAGEQTNGFTDIMDADNSGEIVLKEPKVKVDLRKYTHVHRFFFDEVFD 192

Qy         63 DTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY 122

Db        193 EACDNVDVYNRAARALIDTVFDGGC-ATCFAYGQTGSGKTHTMLGKGP-----EPGLYAL 246

Qy        123 AAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKIL 182

Db        247 AAKDMFDRLT----SDTR-IVVSFYEIYSGKLFDLLNGRRPLRALEDDKGRVNIRGLTEH 301

Qy        183 RVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAG 242

Db        302 CSTSVEDLMTIIDQGSGVRSCGSTGANDTSSRSHAILEIKLKAKRTSKQSGKFTFIDLAG 361

Qy        243 SERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGK 302

Db        362 SERGADTVDCARQTRLEGAEINKSLLALKECIRFLDQNRKHVPFRGSKLTEVLRDSFIGN 421

Qy        303 SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKST-----CINEED-----DTNTER- 351

Db        422 CRTVMIGAVSPSNNNAEHTLNTLRYADRVKELKRNATERRTVCMPDDQEEAFFDTTESRP 481

Qy        352 -----ISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKIN 387

Db        482 PSRRTTTRLSTAAPLFSGSSTAAPALRST-LLSSRSVNTLS 521
```

RESULT 5
T18501
hypothetical protein C0760c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18501
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18501
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3394 <LAW>
A;Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331922; PIDN:CAB11140.1
C;Genetics:
A;Map position: 3
A;Note: C0760c

```
  Query Match             11.0%;  Score 752;  DB 2;  Length 3394;
  Best Local Similarity   21.9%;  Pred. No. 7.4e-24;
  Matches  308;  Conservative  249;  Mismatches  432;  Indels  416;  Gaps   58;

Qy         15 SELEKK------------------KKDSDIITVKNNCTLYIDE----PRYKVDMTKYIE 51

Db       2187 SELEKKNDMLENYKNELKEKNEEIYKLNNDIDMLSNNCKKLKESIMMMEKYKIIMNNNIQ 2246

Qy         52 RHEFIVDKV-------FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYT 104

Db       2247 EKDEIIENLKNKYNNKLDDLINNYSVVDKSI-------------VSCF------------ 2281
```

```
Qy        105 MLGSQPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMV 164
Db       2282 ----------EDSNIMSPSCNDILNVFNNLSKSNKK---VCTNMDICNENMDSISSINNV 2328

Qy        165 AALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDE------------- 211
Db       2329 NNINNVNNINNVNNINNVNNINNVKNIVD-INNYLVNNLQLNKDNDNIIIIKFNILKLFK 2387

Qy        212 -SSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTV-----SQNKQTQTDGANINR 265
Db       2388 LGSCYLYIINRNLKEI----QMLKNQILSLEESIKSLNEFINNLKNENEKNELIKINNFE 2443

Qy        266 SLLALKECIRAMDS---DKNHIPFRDSELTKV-LRDIFVGKSKSIMIANISPTISCC--- 318
Db       2444 EILKLKNNLQDNESCIQNLNNYLKKNEELNKINVKNIF--KYKGYIIHLIQQSNVFCKIF 2501

Qy        319 ----------EQTLNTLRYSSRVKNF-----------KNKSTCINEE---DD------TN 348
Db       2502 KHFNENKIIDQSIINKLLYLKKSFDFYMYDSVIQEIRENKNIIINQDFLTDEYFKHIQTF 2561

Qy        349 TERISILDSKG------SEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILK 402
Db       2562 TKTCNVLIQRGYLSILKDTNNDFFIQN---KQSNQQGNQNGNHINMCNIYPDDEINVTAD 2618

Qy        403 NKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTIN--KKHNNNNNN 460
Db       2619 QQIFD-------GTENVQQSLQNEEDYVNNEEM----YTDKMDLDNNMNGDDDDDDDDDD 2667

Qy        461 NNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLP 520
Db       2668 DDDNNNNNNNNNNNNNNNNMGDEDNHLVNAFNNHNLLTNGNVKSDQINNETLERYEENII- 2726

Qy        521 QPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKN-RSSCDNIMNKKKNNL 579
Db       2727 -QNIYTNDNVDNNQV------IENINKILIKDKQDIINNDELKNEHNNLIRLINESIENA 2779

Qy        580 HLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKM-EDNTPKD--ILY--ESRNVSNMNGN 634
Db       2780 H------------------NLENVYVQNDANNLINDNIKKEETLTYVDEKDNVSN---- 2816

Qy        635 VLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIY 694
Db       2817 -----ESNSKCD--------DDKKENEDIIQAKNENYPVSTHYDNNDDINKDNIN----- 2858

Qy        695 KSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNN 754
Db       2859 --NDNNNDNIND-------------DNNNDNI----------------NNDNNNDNINNDN 2888

Qy        755 NNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNI-NTNKNNQDGNVNYSM 813
Db       2889 INNDNINNDNNNDNNND-------NSNNGFVCELSSNINDFNNILNVNKDNFQG-INKSN 2940

Qy        814 NFCHYNLNDKNY-----LID----LNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNN 864
Db       2941 NF-STNLSEYNYDAYVKIVEAGSALENKKQQD--------------------KKRKYFSD 2979

Qy        865 NNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSS 924
Db       2980 SEITKENGYLGESNSIKIRKVSVGG----DNDNNDDDNNDDNDDDDNDDDNNDDDNNNDD 3035

Qy        925 SSSSS-----------NNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDD 973
Db       3036 DNDDSYDDEEKEQENYSNDVNTELNNHVTFEDISVHNN-----IQENKKNYFSNVHENTF 3090

Qy        974 IIYTINSLNDYMSNTLLHFKEKYTYPTLSTNED-----------IYNKEMEGKHIRLDDQ 1022
Db       3091 NLHVDENVHEDLQN-----YEDYVNNNNNNNDDDNEEESNNSCYIISSDDEGDKKNVSKQ 3145

Qy       1023 DKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNF 1082
Db       3146 NDDDEDDDDNGDDDDNEDDDDNEDDDNGDDD-DNEDDDNGDDD--DNEDDNYDDEDNDNY 3202

Qy       1083 NNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINED 1142
Db       3203 ENEDDDNYANEDDDNYEN------DDDDNYENDDD-------DNYENDDDN----YENDD 3245

Qy       1143 IKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNP 1202
Db       3246 DDN-------------YE---------------NGDDDNDNDHND-----------DNND 3266

Qy       1203 DDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNG 1262
Db       3267 EEKYS--------------------CHDD-------------KNEHTNNDLLNIDHDN- 3291

Qy       1263 DIILLNKKLVQDNIKN----SMDHN 1283
Db       3292 -----NKNNITDELYSTYNVSVSHN 3311
```

RESULT 6
D71606
hypothetical protein PFB0800c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: D71606
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: D71606
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1817 <GAR>
A;Cross-references: GB:AE001418; GB:AE001362; NID:g3845275; PIDN:AAC71949.1; PID:g384528
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0800c

```
  Query Match             10.9%;  Score 748.5;  DB 2;  Length 1817;
  Best Local Similarity   23.6%;  Pred. No. 5.3e-24;
  Matches  356;  Conservative  197;  Mismatches  410;  Indels  543;  Gaps   69;

Qy         79 IIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF-----QYAAGDIFTFLNI 133
Db        391 IFELFTEGCYLFLLPNEKIKSNNIY------------TPNIFLKKLKNYITSNDFV---I 435

Qy        134 YDKDNTKGIFISFYEIYCGK----------LYDLLQKRKMVAALENGKKEVVVKDLK--- 180
Db        436 HINDRTLNRYIKFLSYHCSNNIYVHILFNDLYITLHKKIFI-----NNYDMLIKHYKSTA 490

Qy        181 --ILRVLTKEELILKMIDGVLLRK-IGVNSQNDESSRSHAILNIDLK-----------DI 226
Db        491 DHILYLTNKNE-NLNYLNTILLNNYYSLYNKKENNEKRQSLENLKVKIFPSHLNYKQHTT 549

Qy        227 NKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPF 286
Db        550 NKNVN----------------DPNQQHKHDKDD---------SYDNTYEQMKNNKNKIYP 584

Qy        287 RDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDD 346
Db        585 NEYITTHILQN---------------------------NYEQNLYSFQKK-------DD 609

Qy        347 TNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILK---- 402
Db        610 TNINNIFDL-HKREQIYEYEKEN---ESSDIFRDSYKRKIKEEK-----KKKNIYKYEDH 660

Qy        403 --NKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYN---DNTINKKHNNN 457
Db        661 PLNKEKKKKKKFPFYINYEKGDDKNDNDL-----------YYNNIYSKNLENIQNKNYNNN 709

Qy        458 NNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNN 517
Db        710 NNNNNNNNNNNNNNNNNNNNN-------------NNNNNNNNNNNNNNNNNNNNNFDKY 756

Qy        518 HLPQPNYAFTDTSDFSSL-------DDMNCHLNNNDKSIF-LHKKNLRDNIK-----LKN 564
```

```
Db    757 NIPCTNTNLSLLYDKEKLFLFTYAYDKIQTYTYEELKSKYKISTKIVDRNIKMFLKFLKN 816

Qy    565 RSS-----CDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNT------FFKS----- 608

Db    817 YNNNENTYVDNIIS-KKNIFHLLA-SMKNKVT--------NKTNTHKDIYQFIHSWYHIK 866

Qy    609 --NINK----MEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDEN------HND 656

Db    867 LADQNKEHSFQDDKYLINNLYEKHKVQH---------NTMTHHIIKMEDKKGDIHLMENN 917

Qy    657 NKINNGVINIINNSNVNS--------------------INNS----NMNSINNS-NMNS 690

Db    918 NMLLNNNMSLNNNMSLNNSIPLNNSIPLNNSIPLNNSIPLNNSIPLNNSIPLNNSISLNS 977

Qy    691 ----------------------NSIYKSNYNSNQSISDVQIR-YVNEMDTS--------- 718

Db    978 CISLYNSISLYSNKNTSFNHLYNNIYDTCFIQNNYISNQQVQNYKNEKNTNMEHYNEKKL 1037

Qy    719 ----------------------NKNNDNIFF---------------------------- 727

Db   1038 FIYPIYYLEDKNYFLNVVNNIFFNKNYNNTFFYTCQINILSKGLYYFINYYTLLISSNYK 1097

Qy    728 -DAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQP 786

Db   1098 AEEIKTDDNKC-NINNNNNNNNNNNNNNNN---NNNNNNNNYNNN----NYNNNNVYPL 1149

Qy    787 DNNKNTS----------NIQNINTNK---------------NNQDGNVNYSMNFCH---- 817

Db   1150 INHFTTTFYEMVTYLLKNIYRIHISKFFYIFVALSKFFLMNSYQQSNTNKRENSIHMENV 1209

Qy    818 -------------------YNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQN-----RNDF 853

Db   1210 LYILYIIRKKQYEHVKSILYDKSNENYFRFNENKDIKMENTNMLYNIILNNFSTEDHDEF 1269

Qy    854 EKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKD 913

Db   1270 MTLQKNNEDNNKMIIDNINNVDNIN---------------DLIKSHHCDNNKKED----- 1309

Qy    914 NHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDD 973

Db   1310 ------------TSSLHNKLYNGLHFLIMFLNNYLDNTKHFKI----NHFLSSLFYINK- 1352

Qy    974 IIYTINSLNDYMSNTLLH-----FKEKYTYPTLSTNEDIYN-----KEMEGKHIRLDDQD 1023

Db   1353 -IIPPNMKHMYHLETYLHKNHKIYKNKFFY--------IYNGLDLLKKSYLVHIK----- 1398

Qy   1024 KYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNN--DKNNVDNNNVDNDDDDVDFHNIKN 1081

Db   1399 KLYINSYIKSYNNKKK-------NNNVNGDVYNNFMYKYNI--------YDNIDYIFIKK 1443

Qy   1082 -----FNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAE----KD 1132

Db   1444 KNLFCYTNHLSLLYF-----TYIYS-LNKFYYCTLYYN----ISKCFYYKINIENIHFKN 1493

Qy   1133 NVIKKYINEDIKNMSLEEIDKTAQSIY---EKRKV------LLTKLLLLFKKNVDTQINN 1183

Db   1494 KIILFFIFTQCKYIYIKFFRLLVQSIFSSSEFQKVGKLSLYILSNIILLLVKNSRMKLNI 1553

Qy   1184 ETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCE----SENLRLLYQ 1239

Db   1554 KKKKIIKN-ISKHIYSNN---------------EFINNNKIKKIHTNNNSMSKNLFICNK 1597

Qy   1240 FLVVEYQ---------------------NKSANSVLLNVSSNNGDIILLNKKLVQDNIKN 1278

Db   1598 LLNIQWNYIFPMDLFISSNLSHETELIINKLEQNILNNNNNNNNN----------NNNNN 1647

Qy   1279 SMDHNN 1284

Db   1648 NNKHNN 1653
```

RESULT 7

```
T18277
kinesin heavy chain - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18277
R;deHostos, E.L.; McCaffrey, G.; Vale, R.D.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z18853
A;Accession: T18277
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <DEH>
A;Cross-references: EMBL:U41289; NID:g1526990; PID:g1526991; PIDN:AAB07748.1

  Query Match             10.7%;  Score 732;  DB 2;  Length 1254;
  Best Local Similarity   22.3%;  Pred. No. 1.7e-23;
  Matches  321;  Conservative  217;  Mismatches  398;  Indels  504;  Gaps   59;

Qy      1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYI--DEPRYKVDMTKYIERHEFIVD 58

Db     25 VSSNIRVVCRVRPLTELEKGRNEHSIVHFFDSKSISIRANGP------------QFTFD 71

Qy     59 KVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPG 118

Db     72 RIFGYQETQSQIFEDVAEPIVND-FLDGYHGTIIAYGQTASGKTFTMV-----GDPDSHG 125

Qy    119 IFQYAAGDIFTFLN-IYDKDNTKGIF----ISFYEIYCGKLYDL---------LQKRKMV 164

Db    126 IIPRVIESIFVGISKMREKDTSLSLAFCLKISALELYNEKLYDLYIASKSNLNIREHKQN 185

Qy    165 AALENGKKEVVVKDLK----ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILN 220

Db    186 GIYVEGISEIVITSIEEAYNFLNISNNN-----------RAIASTKMSAASSRSHSVLM 233

Qy    221 IDLKDIN---KNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM 277

Db    234 IELSQQNLSMESSKISKLFLVDLAGSDISHKTGAEGDRMQ-EAKNINLSLSALGKVINAL 292

Qy    278 DSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNK 337

Db    293 TCGVNYVPYRDSKLTRVLQDSLGGNSKTSLIINCSPSNNNEHETITTLQFGTRAKTIKNQ 352

Qy    338 STCINEEDDTNTERISILD-SKGSEMNASSIENVV----IKSNHLLSNNNNNKINRGKIN 392

Db    353 PK-INKKITYHELELFIIKLAKDLEKSRKECEEITRSKNLEINNLLIQLENNQKMVVESN 411

Qy    393 DKIE--RNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTI 450

Db    412 QKLELLNSQISSNHSFD------------NTFKEIENTCENSK--------IIFDD--L 448

Qy    451 NKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNN 510

Db    449 NDHINNNNNNVDENNNTNNNDN---------------------------NNNDNNNNN 478

Qy    511 NNNSHNNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDN 570

Db    479 QYQEESNQYQQEN-------------------NQKDGD--------------QNNSSFDS 505

Qy    571 IMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSN 630

Db    506 I---------------------------------------KVEDLRDLD---DEPDIED 522

Qy    631 MNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNS 690

Db    523 IILNSTLG-------NISDDDDDDDHHSNN-----------NNVDDNNNGEINND---- 560

Qy    691 NSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNN 750

Db    561 -----SDGYLNRSLKDIKIPEISDLNDHNINN-------------------NNNNNNNIN 596

Qy    751 NNNNNN---------------NIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNT--- 792

Db    597 NDNNSNSGGLRVSTSYITSSPNLSIKEYGCH---HNSPPLFSYFKTKDFPPSSDENDKFF 653
```

```
Qy     793 -------SNIQNINTNKNNQDGN-------------VNYSMNFCH------------- 817
Db     654 NDLIAKGENEQQQQQQHNDDDEDIKSTTSNATTTTITTIDMNASHPSGIDDPIEFTIIK 713

Qy     818 ----------------YNLNDKNYLIDLNNKEQ-----------------------KDK 837
Db     714 SDKTITSTIERETIQPSSLSNSTSLLDIETVESSTLPAPPPVTTTTTLTTVTTTKLTKTT 773

Qy     838 NIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMK 897
Db     774 NIPS-NTNDINSIDDFGFSK-----------IEEEGSSSNRKP--------NDTAILSFG 813

Qy     898 NNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHND----NT 953
Db     814 DDDDEENE--DNENEDVIVDSDED----THSGKNNLLN------TFKNDHHRGDFGATPT 861

Qy     954 FTIRRKN------------------------------------NTNINSNIY------- 969
Db     862 KSIFNKNGNITIKEFETPQQQQQQQQQQQQQQQQQQQQQPLILQTTSTNPTIISIKSNKE 921

Qy     970 -----------------QNDDIIYTINSLNDYMSNTLL-----HFKEKYT--------- 997
Db     922 PSPSSSTTTSIKKKNFNKRRSWIIFTIILTITLVSSSLLCLYLPEYKERLVQRRGYLNKL 981

Qy     998 --YPTLSTNEDI----YN-----KEMEGKHIR--LDDQDKYDDNDNNNVDNNNKNNVDNN 1044
Db     982 GIYSDYPTNEKISLAQHNQISLAKELYGGNSKQYYDEMSSFNTAYNHLIEMNHLETAVSK 1041

Qy    1045 VDNNNVDNNVDNNDK-NNVD---------NNNVDND-DDDVDFHNIKNFNNNEYLSYFQ 1092
Db    1042 IFGSAIDLRFSGDDVINSIECKRAIHKLKTNNYVNGDLDQQQQHNY--ITKIDQLSEQS 1099

Qy    1093 KNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYIN--EDIKNMSLEE 1150
Db    1100 KEQNQLIENF--KLDLK---NKTSEIEK---LEKEIKQKDNKIKEKEEKIELIESRVLNE 1151

Qy    1151 IDKTAQSIYE--------KRKVLLTKLL-----------LLFKKNVD-TQINNETSDLRK 1190
Db    1152 -EKGGEKVLEDQIISLRNDKNTLSTQILNLEGDKKSLGVLVIKLNSDKTEIQNEVKELKR 1210
```

RESULT 8
A44259
kinesin-related protein KIF2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C;Accession: A44259; S27871
R;Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
J. Cell Biol. 119, 1287-1296, 1992
A;Title: Kinesin family in murine central nervous system.
A;Reference number: A44259; MUID:93077686; PMID:1447303
A;Accession: A44259
A;Molecule type: mRNA
A;Residues: 1-716 <AIZ>
A;Cross-references: EMBL:D12644; NID:g220467; PIDN:BAA02165.1; PID:g220468
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIP:118908)
C;Superfamily: kinesin-related protein KIF2; kinesin motor domain homology
C;Keywords: ATP; microtubule binding; nucleotide binding; P-loop
F;196-531/Domain: kinesin motor domain homology <KMOT>
F;285-292/Region: nucleotide-binding motif A (P-loop)
F;291/Binding site: ATP (Lys) #status predicted

```
  Query Match             10.6%;  Score 728;  DB 1;  Length 716;
  Best Local Similarity   47.6%;  Pred. No. 1.4e-23;
  Matches  161;  Conservative   61;  Mismatches  102;  Indels   14;  Gaps    8;

Qy       4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db     195 RICVCVRKRPLNKKETQMKDLDVITIPSKDVVMVHEPKQKVDLTRYLENQTFRFDYAFDD 254

Qy      64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
Db     255 SAPNEMVYRFTARPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 313

Qy     123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
Db     314 AARDVFLMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 370

Qy     179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237
Db     371 LQEREVKCVED-VLKLIDIGNSCRTSGQTSANAHSSRSHAVFQIILR--RKGKLHGKFSL 427

Qy     238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297
Db     428 IDLAGNERGADTSSADRQTRLEGAEINKSLLRLKECIRALGRNKPHTPFRASKLTQVLRD 487

Qy     298 IFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNF 334
Db     488 SFIGENSRTCMIATISPGMASCENTLNTLRYANRVKEF 525
```

RESULT 9
T10755
kinesin-related protein 2 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 29-Sep-1999
C;Accession: T10755
R;Sperry, A.O.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z17121
A;Accession: T10755
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-671 <SPE>
A;Cross-references: EMBL:U44979; NID:g2772515; PID:g2772516
A;Experimental source: strain Sprague-Dawley, testes
C;Genetics:
A;Gene: KRP2
C;Superfamily: kinesin-related protein KIF2; kinesin motor domain homology
F;205-542/Domain: kinesin motor domain homology <KMOT>

```
  Query Match             10.6%;  Score 722;  DB 2;  Length 671;
  Best Local Similarity   41.2%;  Pred. No. 2.3e-23;
  Matches  171;  Conservative   78;  Mismatches  138;  Indels   28;  Gaps   11;

Qy       4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db     204 RICVCVRKRPLNKQELAKKEIDVISVPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 263

Qy      64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
Db     264 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKSQNASKGIYAM 322

Qy     123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
Db     323 ASRDVFLLKNQPRYRSLNLE-VYVTFFEIYNGKVFELLNKKAKLRVLEDSKQQVQVVGLQ 381

Qy     181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
Db     382 EYLVTCADD-VIKMINMGSACRTSGQTFANSNSSRSHACFQILLR--AKGRLHGKFSLVD 438

Qy     240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
Db     439 LAGNERGADTSSADRQTRMEGAEINKSLLALKESIRALGQNKAHTPFRESKLTQVLRDSF 498

Qy     300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
Db     499 IGENSRTCMIAMISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------AVQME 549

Qy     359 GSEMNASSIENVVIKSNHLLSNNNNNKIN--RGKINDKIERNNILKNKSFDKPRE 411
Db     550 TEEMDASS-------HGASLTGNEEEELSSQMSSFNEAMTQIRELEERAMEELRE 597
```

```
RESULT 10
D71618
hypothetical protein PFB0285c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: D71618
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: D71618
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1436 <GAR>
A;Cross-references: GB:AE001385; GB:AE001362; NID:g3845143; PIDN:AAC71846.1; PID:g384514
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0285c

  Query Match          10.2%;  Score 698;  DB 2;  Length 1436;
  Best Local Similarity   23.8%;  Pred. No. 5.3e-22;
  Matches  354;  Conservative  224;  Mismatches  524;  Indels  388;  Gaps   71;

Qy         17 LEKKKKD-----SDIITVKNNCTLYIDEPRYKVDMTKYIERHEFI--VDK----VFDDTV 65
Db         47 IDDKKKDLGKSKYDIFTTDSLSTT-TDEVSYSYQIENKNEEKEYLRYYDKQGGIIRQDNN 105

Qy         66 DNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA-A 124
Db        106 NN----ENNNNNICNNDHNNNNICN-----------NENMLTTKKNDNTIINSNIKYLNN 150

Qy        125 GDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALE-----NGKKEV--VVK 177
Db        151 NNIFNTNMVPQKNHTQ-----IFNPYDKSMRNIQLYNKAVSFLKNDGDINSKKNTHDNLM 205

Qy        178 DLKILRVLTKEELIL---------KMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDIN 227
Db        206 FLKNIRSKSNNNLIVNRKITNHVTNNVISGMTNKVIGGMASGMTNNVTSSITNNMTSSMT 265

Qy        228 KNTSLGKIAFIDLAGSERGADTVSQN---KQTQTDGANINRSLLALKECIRAMDSDKNHI 284
Db        266 NNMASGMTSSMTNNMASGMTSSITNNMTSSMTNNMASGMTSSI---------------- 308

Qy        285 PFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEE 344
Db        309 ---TNNMTSSMTNNMASSMTSSMTNNMTSSMT------NNMTSSMTNNMLNNMNRVVTNN 359

Qy        345 DDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGK----INDK--IERN 398
Db        360 IITNMNR-SVSGSKSINMSNLLIINKMDYGNDIYHNNNNNNNNSSSGSNIVSGKYFVNSQ 418

Qy        399 NILKNKSFDKPREGFTSTFGKYSSLNDIDK-------------IKKNKKKGLINYKSTLY 445
Db        419 NSSKNNFFTKVGE---STIRSPTNILDIYKQGNMYMHIPKNADLMNNVSSYSIAHENYIK 475

Qy        446 NDNTINKKH--NNN-----NNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNN 498
Db        476 RDNT-NVTHVLNNNHLVNINNVVNNNNLNNNNNLNNNNNLNS-----------NNNLNSN 523

Qy        499 INVNNNNNNNNNNNNSHNNHL---PQPNYAFTDTSDFS---SLDDMNCHLNNNDKSIFLH 552
Db        524 NNLNNNNNLINNNNLINNNYVRNNQAVNNAHTLNAHFNKSDNVDNMRNHIPNND------ 577

Qy        553 KKNLRDNIKLKNRSSCDN---IMNKKKNNLHLARHSVGSKLTMFSYDPQKNKD-NTFFK- 607
Db        578 NKNIVNMLNLKNMKSINDLSVLINKNKPIHHVINGTEVQQKRSLS-NVQKLKTLNTFPNA 636

Qy        608 ----SNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGV 663
Db        637 KGRFSLINKMA---------SMPNMSTTSSMNMSGLNTSSSEGLTNIINMNNINSVNN-- 685

Qy        664 INIINN----SNVNSINN----SNMNSINNSNMNSNSIYKSNYNSNQSISDV-QIRYVNE 714
Db        686 INSVNNINSVNNINSVNNLNSVNNINSVNNINSVNNINSVNNINSVNNINSVNN 745

Qy        715 MDTSNKNNDNIFFDAISCDNNM-YPNITNNNNNNNNNN-------NNNNNI-------- 758
Db        746 LNSVNNINSVNNINNINYINNINYVNMNKGLNPINNVSNISSLKLLNNNNDIKKKFNTYG 805

Qy        759 --------------------DVENYNNRDGTNNSMKLYAYNS----------HNLF--QP 786
Db        806 KSEASENLSKNVKYIKYIQENIKYLNNLDDNKRKYSLTSINDVGCIKKKKNMNDLFLGKH 865

Qy        787 DNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNI 846
Db        866 DNMLRTDEIPKINLGKNILNNN-----KIINYNDNDKSNII---------------NNV 904

Qy        847 IQNR------NDFE-KKKKTNFYN----NNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDN 895
Db        905 INKNISTDLVNDREGDMNKMNIHNREKDENNYI----NIGDNKIKK---------NQIDV 951

Qy        896 MKNNEMKNNEMKDNEM-----------KDNHIKSNNNNSSSSSSNN----NIYNNINDD 940
Db        952 VNNKVMKLDNMEDEEAMNKLSLISLYPNNNHIINNVNNVNNVNNVNNVNNVNNVNNVNNV 1011

Qy        941 DTFQN-DYCHN-DNTFTIRRKNNTNINSNIYQNDDI--IYTINSLNDYMSNTLLHFKEKY 996
Db       1012 NNVNNVNYMNNMNNVNNVNNMNNVNNMNNVNNVNSVNNIKGINNMNNNNNINMNRSYKM 1071

Qy        997 TYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKN---------------- 1039
Db       1072 NMKKVSKKDNGQNVVSEKRF----SEEKY--NFLKNLIRNNKNMVKLKYLNKFLGKRSGP 1125

Qy       1040 NVDNNVD------NNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQK 1093
Db       1126 SIKNNMNDMMVKMNNNMKDIMHIKDATNI--NKINN--------KLVNLNTNNCISYNSC 1175

Qy       1094 NVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDK 1153
Db       1176 NKMNYIHKCKKKRVLCL---DTKHGKNEI--------KQNEKLIYTNYEIKMFLLNTI-K 1223

Qy       1154 TAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHI-CNNNPDDQFHFYAYS 1212
Db       1224 AIGIVFKKWKFKNFGLYFWY----------------------HIKCIENERD-LNFYI-- 1258

Qy       1213 RLEKDIINLIMLRQIWCESENLR----LLYQFLV----------VEYQNKSA--NSVLLN 1256
Db       1259 KIFNFLFEIITGKNIYYQINDIHNIVALFKEFKIYDCKHVLKKSIKVLNKYAKKNSKEFS 1318

Qy       1257 VSSNNGDIIL-LN---------KKLVQDNIKNS----------MDHNNIH 1286
Db       1319 LFENNQHVVLDINKHMLFNDDEKKLTTCNIKQNEQEQIKTKVLYDHDNIN 1368


RESULT 11
B71605
hypothetical protein PFB0850c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71605
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71605
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-807 <GAR>
A;Cross-references: GB:AE001420; GB:AE001362; NID:g3845287; PIDN:AAC71959.1; PID:g38452
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0850c

  Query Match          10.2%;  Score 696;  DB 2;  Length 807;
```

```
Best Local Similarity   26.6%;  Pred. No. 3.5e-22;
Matches  253;  Conservative  151;  Mismatches  300;  Indels  246;  Gaps   43;

Qy        328 SSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKIN 387
Db         10 SSKEKNEEN--DIINKCDDSN-------KINGKE-NIFAVEKVGINESGHMSNDNINK-N 58

Qy        388 RGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYN- 446
Db         59 QEKNKKKKKKKNTHKKVNINNTHINIHTTNDKNNG-QDINKPEVIERDNIINIKNDTNNI 117

Qy        447 -DNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNN-----MINHMINNNI----- 495
Db        118 LDSSYNEEGNENNRNDINNNNNNNNININNNNINNSCSNNYGLKKKITLLKRNDIKDEGY 177

Qy        496 -NNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKK 554
Db        178 NNENITTLNNKNNLKNNNNYNDNR-----------------------NNNN-------- 205

Qy        555 NLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKME 614
Db        206 NNKNNINNNNNNNC------------------CSEKTLEQREKEYNKIRARIFSNFNKKQ 247

Qy        615 DNTPKDILYESRNVSN--MNGNVLLGLNKNTHHDISTKDENHNDNK---INNGVINIINN 669
Db        248 KNVQKT---EQNNLNHTYLNNNIINNI------------NNGDNQYAYINN-FYHIYHN 290

Qy        670 SNVNSINNSNMNSINNSNMNSNSIYKSN----YNSNQSISDVQIRYVNEM-DTSNKNNDN 724
Db        291 NSYNHIYRQNNIPICNINNHAPNIEKLNNPYYYHDN------HIAYTNYMYSTQNKM--- 341

Qy        725 IFFDAISCDNNM---------YPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKL 775
Db        342 ---------NNMKTKQIGHYGINNEDNNNNNNNNINNNNNNNINNNNINN---NNVPLCI 389

Qy        776 YAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLID-LNNKE- 833
Db        390 PQLDNYNKTKNNFNQGTNNFNQGTNNFNKCTNNFNNAKNHIKHNINNTNKNIEHLNNHSI 449

Qy        834 -----QKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCG 888
Db        450 YNFVYPENKNIYDANGNLINNNISYTQLKMNNNINFNIHMESPINQQHNNTFKV-----N 504

Qy        889 SHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNN--IYNNINDDDTFQND 946
Db        505 NDTNFFNEPTNKMKK---KNKEKKNIHFNNNNNN-------NNNKCLYKDIN-----QND 549

Qy        947 YCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNED 1006
Db        550 H-----------NNSIINTN--QNFDHINNV-------KNTEQNLQKKHNKMSQVSKQS 588

Qy       1007 IYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNV---DNNNVDNNVDN------- 1056
Db        589 NNKNNKNNSHLK----KQININTNNNMDNKNNSHISKNVIVDDNKLKSSHADNSNEIVTK 644

Qy       1057 ---------------NDKNNVDN----NNVDNDDDDVDFHNIKNFNNNEY-----LSYF 1091
Db        645 GKKKKNTNKKKKINNINSVNNVNNINSMNNINSMNNIISMNNVNNMNNPMYFPNVNIQKD 704

Qy       1092 QKNVDTIINNCLNSLDISSM---YDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSL 1148
Db        705 DSNIALLYNNKPN-IDFNNFQLNHINNHMIQNNIMTNNVML-NNNLTTSNFNYNLINYSY 762

Qy       1149 E---------EIDKTAQ-SIYEKRKVLLTKLLLLFKKNVDTQINNETSDL 1188
Db        763 EPFYEENLMNDLDYCRDISLYEKR----------YDRGDNLQQNHKRYDI 802
```

RESULT 12
S71629
sensory transduction histidine kinase dhkA - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 24-Sep-1998
C;Accession: S71629
R;Wang, N.; Shaulsky, G.; Escalante, R.; Loomis, W.F.
EMBO J. 15, 3890-3898, 1996
A;Title: A two-component histidine kinase gene that functions in Dictyostelium developm
A;Reference number: S71629; MUID:96324397; PMID:8670894
A;Accession: S71629
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-2150 <WAN>
A;Cross-references: EMBL:U42597
A;Experimental source: strain Ax4
C;Genetics:
A;Gene: dhkA
A;Map position: 6
C;Superfamily: response regulator homology
C;Keywords: autophosphorylation; phosphoprotein; phosphotransferase; two-component regu
F;2027-2142/Domain: response regulator homology <RRH>
F;2076/Binding site: phosphate (Asp) (covalent) #status predicted

```
  Query Match             9.9%;  Score 676;  DB 2;  Length 2150;
  Best Local Similarity   24.5%;  Pred. No. 6.8e-21;
  Matches  242;  Conservative  146;  Mismatches  255;  Indels  344;  Gaps   41;

Qy        416 TFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNN 475
Db         85 SYGNHSFIHNVSPTSP-------SYDINNNNNNNNNNNNNNNNNNNNSNNNNNNNNN 137

Qy        476 N--------------------------------------------NDSSSM--VNNMINH 489
Db        138 NKNNYNNNYYYSPIENSNISKSLEESVLNQFPHNFNLNSSNNNYLNNSSSLHNINQSVNS 197

Qy        490 MINNNINNN----INVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNN 545
Db        198 LSNNNNNQTNQQPINNNNNNNNNNNNNNSNNSN-------------------NSNNNNN 237

Qy        546 DKSIFLHKKNLRDN-IKLKNRSSCDNIMNKKKNNLHLARHSVGSKL---TMFSYDPQKNK 601
Db        238 GN----NNNNITDSPTKSKRHSTYETNIGS-----HQRRKSIQSLIANSAIHSFSKLKNK 288

Qy        602 DNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINN 661
Db        289 ----------PLSSSTPSTV-----------NTCGAVNNNSNN-------NNNNNNST 319

Qy        662 GVINIIN-----NSNVNSI-------NNS--------------------NMNSINNS-NM 688
Db        320 GSLGAIPMDRSFDGNINTITEESTGGNNSPRSNCGSNCGSNGGIPLSPRNLSSLNSGVNV 379

Qy        689 NSNSIYKSNYNSNQS-ISDVQIRYVN-EMDTSNKNNDNIFFDAISCDNNMYPNITNNNNN 746
Db        380 SPRNIHLNNLNNNSSNLPPLSPRHINFHINVSNLNNNN--------NNNINPNNNPNNSN 431

Qy        747 NNNNN---NNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKN 803
Db        432 NSNNNVSPRNNNHNISPRGSNISPRSNN-------GGSTTISPRNISNNNNIIN------ 478

Qy        804 NQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCD---------NNIIQNRNDFE 854
Db        479 ---------------NINNNNILTPPRNSPRLE-NVNPTNSPRLLATSLNSTLPIVSSLT 522

Qy        855 KKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDN 914
Db        523 SSNNNNQSNNNTNPSINNNNGRN-------GHCIQTISEEILGNKPVVYNNGNNN----- 570

Qy        915 HIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQ---- 970
Db        571 ----NNNNTNNSTTSNNNITTNNNNN------------------NNNNINNNVLSTPRK 607

Qy        971 ----NDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYD 1026
Db        608 RTKGNHS---KTNSLQDFETS------------SMNGGDDSISGAGSGGSLRRRNKDDND 652

Qy       1027 DND------------NNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDD 1073
```

```
Db        653 ENDGNSNTNSNNSNNNNNNNNNSSNNNNNNSNNNNNNNNNNNNNNNNNNNNNNNNNNN 712

Qy       1074 VDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDN 1133
                : :|  | |||            || |   :   :        |:  :    : | |
Db        713 NNNNNNNNNNNN------------NNYHNGATMMMSH-------NHQSIGMSSSPKKN 752

Qy       1134 VIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLV 1193
               | :      :| ||  : : | :|             :| ||       |   :  :   ||
Db        753 NFKPF----SRNCSLMGMGRRAWAI----------ILGLF-------IVGSSISILATLV 791

Qy       1194 MCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNK----S 1249
               : :    |:  | |      :|: :|     |||                 :|: |:
Db        792 LRYSEENSIADDF-----ARVARD--RFTMLR----------------IEFNNRLYITQ 827

Qy       1250 ANSVLLNV-SSNNGDIILLNKKLVQDN 1275
               |:||:|   |.: |   :    ||  ||
Db        828 TLSLLLSVFPSTSEDQFVPFSKLWSDN 854
```

RESULT 13
T23607
hypothetical protein K11D9.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23607
R;Wall, M.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19769
A;Accession: T23607
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-747 <WIL>
A;Cross-references: EMBL:Z92807; PIDN:CAB07264.1; GSPDB:GN00021; CESP:K11D9.1
A;Experimental source: clone K11D9
C;Genetics:
A;Gene: CESP:K11D9.1
A;Map position: 3
A;Introns: 57/1; 96/3; 169/1; 388/2; 542/2; 646/2

```
  Query Match             9.6%;  Score 656.5;  DB 2;  Length 747;
  Best Local Similarity   37.5%;  Pred. No. 1.4e-20;
  Matches  157;  Conservative   80;  Mismatches  153;  Indels   29;  Gaps    9;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFD 62
              ::|  | ||||||:: |    |  | :|:||: :   : :|: :||:|||:: :|    |  ||
Db        319 NRISVCVRKRPLNKKELTKAEVDVITIPSRDITILHQPQTRVDLTKYLDNQKFRFDYSFD 378

Qy         63 DTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQ 121
               :  :|  ||   |  ||:  :::||    :|||||||||||:|| |        |: : ||:
Db        379 EYANNELVYRFTAAPLVKTVFDNG-TATCFAYGQTGSGKTHTMGGDFSGKKQNASMGIYA 437

Qy        122 YAAGDIFTFL---NIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
               | |:|  |        ||   :   :|:|||  |  ||||  :   :   ||:  ::|  |
Db        438 LTARDVFRMLEQPQYRRKD--LSVHCAFFEIYGTKTYDLLNDKAELRVLEDKMQKVQVVG 495

Qy        179 LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFI 238
              ||  :    :::::  :   | |:|   |   |  |   ||||||  | |:    |    || : |
Db        496 LKEEQAHNEQDVLELINKGTLVRTAGTTSANANSSRSHAIFQIILRQGKK--VWGKFSLI 553

Qy        239 DLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDI 298
              ||||:|||  ||    :: |  :||||| |||||||| |       : :|:|||  ||:|| ||||
Db        554 DLAGNERGQDTRECDRDTRKEGANINTSLLALKECIRGMARNSSHVPFRQSKLTMVLRDS 613

Qy        299 FVG-KSKSIMIANISPTISCCEQTLNTLRYSSRVKNF----KNKSTCINEED-------D 346
              |:| ||:::||: ||| ||  :  ||||||: |||         ::| | :|:            |
Db        614 FIGEKSRTVMISMISPGISSSDHTLNTLRYADRVKEMGTDGSGEATPIRDEELFLPPSAD 673

Qy        347 TNTERISILDSKGSE--------MNASSIENVVIKSNHLLSNNNNNKINRGKINDKIER 397
               : |    :    |               |   :   :|:   :: :|   :   :   :  |:::
Db        674 KSDEEYDEMVEKQEHRRVAVDHVRNLKDMSEKIIRETTIVLSNEPSAAQKAECLAKLDQ 732
```

RESULT 14

S14556
asparagine-rich protein (clone 18C1) - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 01-Dec-2000
C;Accession: S14556; S14548
R;Schreiber, L.; Deutsche, U.; Storck, T.; Mueller-Hill, D.
submitted to the EMBL Data Library, December 1989
A;Reference number: S14469
A;Accession: S14556
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1256 <SCH>
A;Cross-references: EMBL:X17485; NID:g9846; PID:g1335715
A;Accession: S14548
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 'Q',210-1251,'NEENDETALDTIIQPPIKKKK' <SC2>
A;Cross-references: EMBL:X17488; NID:g9852; PID:g9853

```
  Query Match             9.4%;  Score 646;  DB 2;  Length 1256;
  Best Local Similarity   22.1%;  Pred. No. 6.9e-20;
  Matches  264;  Conservative  149;  Mismatches  341;  Indels  440;  Gaps   44;

Qy        336 NKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKI 395
              | :   || |: |           :|:   :   |  :   |  :   :||: || ||      :  |
Db        150 NNNYLINNMDNNN------IDANNNNNNNNNNNNINNLSHLIHNNMNNN------STLI 197

Qy        396 ERNNILKNKSFDKPREGFT-----STFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNT- 449
               ||:::|::    |  |    :| |    |   |   :: | ||             :||||
Db        198 NRNHLIKSERIKKGDVGTTTWTKSTTEGHPEYLPRIPGVRFNPKK---QQWLAAWNDNTR 254

Qy        450 ------------------------------------------------------------ 449

Db        255 EIRRYFSVKQYGFEQARILAVKARQEAEKAGARCKPMFHVHGSRKAVDAAITNDLLRSEM 314

Qy        450 ------INKKHNNNNNNNNDNNDNNNDNNNNNNNNDSSSMVNNM---------------I 487
                    ::  |||||||||:||  :::  |||:|  :  ::   ::||                  :
Db        315 EENFNNMHMNHNNNNNNNNNNNQHHQNNNSNIHHNNHHMHNMNLNINPNHLGHNMSHNM 374

Qy        488 NHMINNN-----INNNINVNNNNNNNNNNNNSHNN-----------HLPQ---------- 521
              ||  |||      |||||| ||||||||||:||:||                |  :
Db        375 NHNYNNNNNNMYINNNINNNNNNNNNNNNSNSNNNNNNNNIGNDVIHRKETVKVEIKGVK 434

Qy        522 ------PNYAFTDTSDFSSLDDMNCHL--NNNDKSIFLHKKNLRDNIKLKNRSSCDNIM- 572
                    :: :|   |||||   |  ||||    |      |        | |
Db        435 RGRGRPPKRKLSEESQM-SLDDMEQTLCRNNNDNMELLECMEAYD-------KDCSRPMK 486

Qy        573 ----NKKKNNLHLARHSVGSKLTMFSYDPQK-----------------------NKDNT 604
                  | :| :  ||  |:|   |  :  :|                              :: |
Db        487 GVSYNDRKGS-WLAYWSIGKNFQMRRFPIKKLGFEKAKELAIQCRLEAEQAGATTTENRT 545

Qy        605 FFKSNINKMEDNTPKDILYESRNVSN--------------MNGNVLLG----------- 638
                 |:  :         ::: :  :: |                       ||||: :
Db        546 KRMRNLLTLNSENALEMMIDQNSIDNDDSNIHQHMKEKGGGMNGNLRMSRMQNYAGAGIS 605

Qy        639 ------LNKNTHHDISTKDEN--------------------------------------- 653
                     || :    |  ||
Db        606 GQSVHQTNKLLNSDAGQDSENEFSPTKRTRAPRGRRMESLTARASALTPVEGVRFDPYSY 665

Qy        654 -------HNDNK-------------------------------------------INNGV 663
                      |:|                                                  |  |
Db        666 SWFAKYLENENSKEPKISKYLLKKWGFNKAHSLAVHTVKCAYKAVPFTDEELVNIFNVDV 725

Qy        664 INIINNSN-----------VNSINNSNM---------NSINNSNMNSNSIYKSNYNSNQ 702
               |:||| |             | : |: |          | : |   |   :   :| :||
Db        726 NNLINNQNNMINLGFHDGFGTNGLMNNQMINQMKTQLNNQMTNQMTNQMTNQMTNQMTNQ 785

Qy        703 SISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNIDVEN 762
               : : :  |:|      |  |           | |  :||  ||  |.  |  |  : |
Db        786 MTNQMTNQMTNQMTNQMTNQVN--------NQMTNQMTNQVNNQMTNQMTNQVNNQMTN 836

Qy        763 YNNRDGTNNSMKLYAYNSH----------NLFQPDNNKNTSNIQNINTNKN----NQDG 807
                  || ||     |||                 |:|  :|:  |: :  |:|   |    | |
```

```
Db        837 QMTNQ-MNNQMK-NQVNSHINSQVAGHMNANMFNGNNSNGTNMLGNMNMLNNMVNSNMDN 894

Qy        808 NV--NYSMNF---CHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKK---KKT 859

Db        895 NVNNNYSNNYPGSGNNGLNNMTSCVMNNNGAEIDMD----DNDNIEN-NDHENMMTLEKE 949

Qy        860 NFYNNNNIV----IVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNH 915

Db        950 DSINNDSIITTTATNINSCNNNS-----------NNANSLLNRLVMSNGSNGN-TKLNI 997

Qy        916 IKSNNNNSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNI----NSNIYQN 971

Db        998 INNYNNNNNNNSSSGGGVMVNINN--------VHNNDNMNTYNNNNNNILNDGNSVVGNN 1049

Qy        972 DDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNN 1031

Db       1050 ESMIHKKNVVN-----------------------CINNDVE------DDLGNMNNNHNN 1079

Qy       1032 NVDNNNKNNVDNNVDNNNVDNNVDNNDKN-NVDNNNVDNDDDDVDFH-NIKNFNNNEYLS 1089

Db       1080 NNNNNNNNNNNNNNNNNNNNNSNNNNNNNNNLNVFMNEHNNPTNNRGIYGSNESINNNGLMD 1139

Qy       1090 YFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSLE 1149

Db       1140 EGRLN---FIHTAGKHVDKNNMQSGSTTVVNVL---------NNDENYIKSNINYTNMD 1186

Qy       1150 EIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPD 1203

Db       1187 NMKNNDNPIDDDNPNNIVEMNLRVEKNI---LNNNNDE--------HIINNTED 1229
```

RESULT 15
T18429
hypothetical protein C0345w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18429
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18429
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1711 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325381; PIDN:CAB11106.1
C;Genetics:
A;Note: PFC0345w

```
  Query Match             8.9%;  Score 611;  DB 2;  Length 1711;
  Best Local Similarity   23.8%;  Pred. No. 2.8e-18;
  Matches  284;  Conservative  202;  Mismatches  406;  Indels  300;  Gaps   57;

Qy        127 IFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRVLT 186

Db        665 ILPFTDFY-----MNIEIS-YQIKC--------KRKDSLHSKWEKRTVSVRSI-YLCIHT 709

Qy        187 KEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERG 246

Db        710 HSRYI----------KRSNKYQNDE---------FDENVSHKNDAVGSIAY-EMEQNE-- 747

Qy        247 ADTVSQNKQTQTDGANINRSLLALKEC-IRAMDSDKNHIPFRDSELTKVLRDIFVGKSKS 305

Db        748 -----INEQERRDG-----EMLGVDEMENRNKIENYEHIDNASSEISN-KENCLIQKNMS 796

Qy        306 IMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILD----SKGSE 361

Db        797 NNLSNDIESIICLSNV--------EIKMF----------NDPNITNITILEKRNYSYGIE 838

Qy        362 MNASSIENVVIKSNHLLSNNNNN--------------------------KINRGKINDK 394

Db        839 LTEYNDKKDLV-GYWLLSQNNEKDMKELYHILCAIKKKNPKAARIPSFYPKINMNNSMFN 897

Qy        395 IERNN---ILKNKSFDKPREGF---TSTF------GKY--SSLNDIDKIKKNKKKGLINY 440

Db        898 YHENNISTIYKNFSANLIEPSYFINTSEHEKDERDGKYLEASINDYMSDDKKKKR----Y 953

Qy        441 KS--------TLYNDNTINKKHNNN-----NNNNNDNNDNNNDNNNNNNNDSSSMVNNMI 487

Db        954 DSIESLRGSDKIKNDQIYQGGHSSSLLYYYDNNNDDNNNNMYDSSSSSNHNYYILTNDKR 1013

Qy        488 NHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDK 547

Db       1014 LNM-DNFINNNLEINNSQNKVIEKNLEYIN-----NVKLTKTSNYEQSNNTN---SKDEH 1064

Qy        548 SIFLHKKNLRDNIKLKNRSSCD-------NIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600

Db       1065 NISSDKSKKEDTLNLSRKSSYEYNNKILQSTSNKSLNGAY---------ENNLFSGKKKKN 1116

Qy        601 KDNTFFK--SNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKN-----THHDISTKDEN 653

Db       1117 K-GTVLKDIEHINDIQDKYPEDL-----NINCVNKYVIENEEKHLLPLELEYNLVSSDEK 1170

Qy        654 HNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYN----SNQSISDVQI 709

Db       1171 FGLNKIK-------NDNNIIYMKHQNYHNLYDDNQKKHILFDTNKNVSIQRNNNINSVIK 1223

Qy        710 RYVNEMDTSNK---NNDN--------IFFDAISCDNNMYPN---ITNNNNNNNNNNNNN 755

Db       1224 TNHYEVEKNNKDQRNYDNFTCDKKKKIYYNIINSDKDIYHNNIIYTKNEKEGIGNIHLNR 1283

Qy        756 NNIDVENYN--NRDGTNNSMKLYA---YNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVN 810

Db       1284 NDKDITNFELLKLDGVKEFLDTFKDSYIDCHN--------KKENILNM-TNKNKEDHQI- 1333

Qy        811 YSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIV 870

Db       1334 --IDVADKIFNETNMITMDNNKIYDDKNVH------------EKKCTHNDVIHHNMDIL 1378

Qy        871 NNNMGNNNSPRMKYGLCGSHTSIDN-MKNNEM------KNNEMKDNEMKDNHIKSNNNNS 923

Db       1379 STSIKNNE----------ENLFIDTYQKQNRIGDIYMNRINILQEDDDDDNHNNNHNNNN 1428

Qy        924 SSS-------SSSNNNIYNNINDDDTFQ--NDYCHNDNTFTIRRKNNTNINSNIYQNDDI 974

Db       1429 NNNKLILFEYTKNDQMLHNNKNNLEGTEEFSDFIEKKNKIKIKNKNES------YHKIDE 1482

Qy        975 IYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVD 1034

Db       1483 SLLSNEKNNKVSLLLIN----------------NNKDSSSVDNNKNNNNKNNNNKNNNNE 1526

Qy       1035 NNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNN--------E 1086

Db       1527 NNNKNNKNNNNDSFSKDNNLINNDNNNNNNNNDSFSKDNNLINNDNNNNNNNNNKVIKKE 1586

Qy       1087 YLSYFQKN--------------VDTIINNCLNSLDISSMYDD-TKEILNNILLSKYKAEK 1131

Db       1587 IIDDKEKNDIHKRDNIYIKDVSVSPLINNHPN---LNSMRKDRTIEPL-KIINGKNKLIK 1642

Qy       1132 D-NVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLF--KKNVDTQ 1180

Db       1643 DLKKIQEQVERKIRKYKIQMDQENKKPPPSKNKINMKSINLDIDDDQNVDSQ 1694
```

RESULT 16
T18477
hypothetical protein C0485w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2002
C;Accession: T18477
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z18937
A;Accession: T18477
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2523 <LAW>

```
A;Cross-references: EMBL:AL008970; NID:e1407852; PIDN:CAA15620.1
C;Genetics:
A;Map position: 3
A;Introns: 148/3
A;Note: C0485w

  Query Match             8.8%;  Score 603.5;  DB 2;  Length 2523;
  Best Local Similarity   21.8%;  Pred. No. 8.6e-18;
  Matches  360;  Conservative  231;  Mismatches  548;  Indels  513;  Gaps   78;

Qy          1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDK- 59
Db        772 LNSYQKGIYKLIELRTNEKKIKNLDITSDVNTKDTSLNVLLTKEMVNKKKEGPPFFNNKG 831

Qy         60 -----VFDDTVDNFTVYENTIKPL--IIDLYE-------------------NGCVCSCFA 93
Db        832 EMCLENHEDMVDIFG--EGKMKGIKNVVDTYDKHNNDNNNNNDSSSSNKCCNNCCSSSYD 889

Qy         94 YGQTGSGKT-------YTMLGSQPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISF 146
Db        890 KGKEKKKQTTEINILINSMCDINNYTHTNTHSVIIKEDYDKIQQNKISSKNDTFNEYSSF 949

Qy        147 Y-------EIYCGKLYDLLQKRKM----VAALENGKKEVVVKDLKILRVLTKEELILKMI 195
Db        950 VFSLNMNTQILKNKLLEMKKKNDLDMYGCNEILKGENEIGMDPLMKIDQTNK---IVSKV 1006

Qy        196 DGVLLRKI-GVNSQNDESSRSHAI--LNIDLKD-INKNTSLGKIAFIDLAGSERGADTVS 251
Db       1007 DGSNFNKVDGINFNKTDGSNFNKIDGINFNKTDGINFNKT----------------DGIN 1050

Qy        252 QNKQTQTDGANINR----SLLALK-ECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSI 306
Db       1051 FNK---TDGINFNKVDDNIFNKIKDEVEKYVDPLPSHI--RTDDMRKKKSELLLSKDGSI 1105

Qy        307 MIANISPT-------------ISCCEQTLNTLRYSSRV-----KNFKNKSTCINEEDDTN 348
Db       1106 IISNLDTSHFEINLSRSEIQNEMCKENSFVKCQLENKLILELEKEIKDEEKNLQNELERS 1165

Qy        349 TERISILD-SKGSEMNASSIENVVIKSNHLLSNNNNNKI-----NR-GKINDKIERNNIL 401
Db       1166 NWSIDIEDLDKDLIINKESRD---IKYKHWIDINKDNYMMIYQDNKCGRRKKMISQNKLL 1222

Qy        402 ------KNKSFDKPREGFTSTFGKYSSLNDI-DKIKKNKKKGLINYKSTLYNDNTINKKH 454
Db       1223 IKKKRIKMRNHEKKRK--IRFFFKLYKRNDTHKKLRPIRFVRHVDVKLDNLNDKTVMLKN 1280

Qy        455 NNNNNNNNDNNDN------NNDNN--NNNNNDSSSMVNNMINHMINN-----NINNNINV 501
Db       1281 EIRDVKGEDKGEDVYFDFLNKDNNMGNMENKKNVKNVKNVKN--VNNVKDVKNVNNVKNV 1338

Qy        502 NNNNNNNN------------------------NNN----NSHNNHLPQPNY-------- 524
Db       1339 NNVNNVNNVKDVKNMEHIDKYNKKEVMIKKKGESNNVPHKEKHNNKKNYCNYDLGMHSLQ 1398

Qy        525 -AFTDTSDFSS---LDDMNCHLNNNDKSIFLHKKN----------LRDNIKLKNRSSCDN 570
Db       1399 NRHTITSEVSSKFLCKNMKNYFDKSNNSIEIHKISASNIFRHTMCVASNIKGENKNNGNN 1458

Qy        571 IMNKKK------NNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMED-NTPKDILY 623
Db       1459 INYKGPATKALVNKLFISKKESKRAIT-----SSKKRDD----DNINVIKKINTPSQKVS 1509

Qy        624 ESRNVSNMNGNVLLGLNKNTHHD-ISTKD-----------ENHNDNKINNGVINIINNS 670
Db       1510 EKRNNNNNNNVLGDKNKNKNNDELFTKEIKKSTISKQKKGKNEGNTKTHKDNINILNED 1569

Qy        671 -----------NVNSINN-SNMNSINNSNMNSNSIYKSNYNSNQS--ISDVQIRYVNEMD 716
Db       1570 VDHFKQPSLRLEVTKKNNKNNKNNKNNKKFNDN--YNNNHNNNNSNDFEEYKEEHIATNE 1627

Qy        717 TSNKNNDNIFFDAISCDNNMYPN--ITN----NNNNNNNNNNNNNNIDVENYNNRDGTN 770
Db       1628 IVRESESDLY---VSCDEGCYKNGDIYNMEIINNVDNIHKIDKKGNDID---YKDKSLEN 1681

Qy        771 N-----------------------------------------SMKLYAYNSHNLFQPDN 788
Db       1682 NKKKQMKGLIKILPSSYTEKEKEKEKEKERKKKKNIYIPLTIASRKTVAYPINDTTKDVK 1741

Qy        789 NK------NT----SNIQNINTNKNNQDGNVNYSMNFCHYNL------------------ 820
Db       1742 NKLHTLKRNTCVTYCNVDNIQNKKKKGDDKKNIKRD-QHVGLEKFLDEMSAMFEKKKKIK 1800

Qy        821 -NDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKK----------KKTNFYNNNNIVI 869
Db       1801 KDDINKKEDINKKDDINKKDNINKKNDINKKDDINKKGDINKKDDINKKNN--- 1857

Qy        870 VNNNMGNNNSPRMKYGLCGSHTSIDNMKNN-EMKNNEMKDNEMKDN------HIKSNNNN 922
Db       1858 YNNNSNNNNVVKK---FSKTHQNEEKIKGNITVIRNKLKDKGKKENIGLKKKKIERKNTT 1914

Qy        923 SSSSSSSNNNI--------YNNINDDDTFQNDYCHNDNTFTIRRKNNTNIN--------- 965
Db       1915 TIATKKHDNIIDIKKKNEKENKMIKNSKFQ---CLKNKGTQIENKKNMIISGGKKTNQLV 1971

Qy        966 --SNIYQNDDIIYTINSLNDYMSNTLLH-----------------------FKEKYTYPT 1000
Db       1972 RMKNSKHQNKVEYNLNNISKELEKKKIYMRYFKKGIKDKIENMGNLKVSRDMKKKKKSDL 2031

Qy       1001 LSTNE-----DIYNKEM---EGKH-IRLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVD 1051
Db       2032 IKNKEGGLLYEWYNKYIYNGDGKEKLKMDDSKNYKD----NICDKNKNNMCDKSKNNMCD 2087

Qy       1052 NNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSM 1111
Db       2088 KN-----KNNMCDDNKNNMCD-------KSKNN-------------------------M 2109

Qy       1112 YDDTK---EILNNILLSKYKAEKDNVIKKY--INEDIKNMSLEEIDKTAQSIYEKRKVLL 1166
Db       2110 CDDNKKEIEIVNIIHTS-------NARQNYNKTNEKIKNISREYMTESYGDISTP----- 2157

Qy       1167 TKLLLLFKKNVDTQINNE-----------TSDLRKDLVMCHICNNNPDDQFHFYAYSRL 1214
Db       2158 ------LKKNVNVNVNKDIHIINDKHERICTKNNNPNLHICQPTNDQDGDK--------- 2202

Qy       1215 EKDIINLIMLRQ-------IWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILL 1267
Db       2203 KKNIVKFKSVENRTPHTYLLFKNDENKAYLEMVKSVNYMNK------------------ 2243

Qy       1268 NKKLVQDNIKNSMD-----------HNNIHKK 1288
Db       2244 -KKGAQHNITNKIDVENSSPVHVPPHNIGYKK 2274

RESULT 17
T18489
hypothetical protein C0820w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18489
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4981 &lt;LAW&gt;
A;Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331910; PIDN:CAB11128.1
C;Genetics:
A;Map position: 3
A;Note: C0820w

  Query Match             8.8%;  Score 601.5;  DB 2;  Length 4981;
  Best Local Similarity   20.3%;  Pred. No. 2.1e-17;
  Matches  313;  Conservative  256;  Mismatches  457;  Indels  519;  Gaps   75;

Qy         17 LEKKKKDSDI-----ITVKNNCTLYIDEPRYK----VDMTKYIERHEFIVDK----VFDD 63
```

```
Db     1655 LNKKKNDSQVDFNNFIILSN---MYNNSDTFKKKNIQNVLQLYLNDYQQVYKKNVVHIFNM 1711

Qy       64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123

Db     1712 YISNVVTHKE---------MENG------MYRKEGNKR---------------------- 1734

Qy      124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRK------MVAALEN------GK 171

Db     1735 ENNTYDKINIDDKNNFG------QPIY--KYVDKLTSKRNTLDNSSINCLESSYVTKGSN 1786

Qy      172 KEVVVKDLKIL--RVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKN 229

Db     1787 TKIRIKDIEKIGNKLDTSNQLVMKSLFFLFEDKLG---------SYILQNDMNFSITKN 1836

Qy      230 TS----------------------LGKIAFIDLAGSERGADTVSQNKQTQTDGANI-NR 265

Db     1837 FKQEIASYYIQSLLYEKNFFQVLFLLSNIYYSDVANKKIGAFSLFGKSSNVIHAENIRNT 1896

Qy      266 SLLALKECIRAMDSDK-NHI-PFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLN 323

Db     1897 LILSMGLCMKKINSHKFNHLYNIREAYVSKEIFHIYLEKI-----------------YLN 1939

Qy      324 TLRYSSRVKNF---KNKSTCINEEDDTNTERISILDSKGSEMNAS---------SIENVV 371

Db     1940 IAKENENFENILFDKNDSVGIIPTEIIN----KIIGNKQKYVKTNYRDQINLYIKIKNLL 1995

Qy      372 IKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKK 431

Db     1996 LSSIH-ISNSNTSDLN--KIN--------VEGQKYEDP------------------VKR 2025

Qy      432 NKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNNDSSSMVNNMINHMI 491

Db     2026 TLSIMRIITNNSIRNEDMDNNKNNSNNNKNNNNKNNSNNNDNNNNKQSN-------YHNS 2078

Qy      492 NNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNND---KS 548

Db     2079 HDGIKKGVGYKNNSSSNTNDN--------------------KIDLSNMHNNNNNIVMNS 2117

Qy      549 IFLHKKNLRDNIKLKNRSSC-DNIMNKKKNNL-HLARHSVGSKLTMF----SYDPQKNKD 602

Db     2118 DYIINNGFTSNGTCSKSPNCNDKLMSSQKEEIAHIYKYMSKEKLHFFLLQLLVDPIL--- 2174

Qy      603 NTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLG--LNKNTHHDISTKDENHNDNKIN 660

Db     2175 GSLYQEKESVLQSSLLKSLRMISKRLKRIKGIILENDKYNNTKKEDLIKNNRILNDESKN 2234

Qy      661 N--------GVINIINNSNVNSI--------NNSNMNSINNSN---------MNSNSIYK 695

Db     2235 NMFEKNYSCEILNYIDINNKRCILKDICEDYLKENLSNVKNYSCLCIYKFIIIYSSLYYL 2294

Qy      696 SN---------YNSNQSISDVQIR--------------------YVNE-------MDTSNKN 721

Db     2295 SNIDNLSFPIHNTTDDISNVSLEKFVLYFKFFYRGIINKSEYYFKEEIIKKYGEDGQYNN 2354

Qy      722 NDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSH 781

Db     2355 DDYMSINNNTKNNNHYSNNNHHNKNNSNNNNSNNNN----NNNNNNNNNNN-----YN-- 2403

Qy      782 NLFQPDNNK-NTSNIQNINTNKN-----NQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQK 835

Db     2404 -----NNNKGNINKVHNKTIRKNIDRLHQLNSNINLEKEFYYIKL-----IVEL-DKKMN 2452

Qy      836 DKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSID- 894

Db     2453 IKHLNKSSNNFI--RNDY----------FSELLIQHNGNSNNIEETKTYKNDYHYHPVDI 2500

Qy      895 NMKNNEMKNNEMKD---------------------------------NEMKDNHIKSNNN- 921

Db     2501 NKESYEILLHSLKTIYNSLFIRHPLNFQYIFIIVDVLIGLFLKIVFNEKKRIYQEFQKNR 2560

Qy      922 ----NSSSSSSSNN-----NIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQND 972

Db     2561 EGCCNKEESGHENNSKGIKNIVSKRGEENMGE------------KKRNNKRTKYKVQEEK 2608

Qy      973 DIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNN 1032

Db     2609 E-------------NVTL---DMYKYKIIKNDHDIYN-----------DEHKCNDNNNDN 2641

Qy     1033 VDNNNKNNVDNNVDNNNVDNNVDNNDK---NNVDNNNVDND---------DDDVDFHNIK 1080

Db     2642 -DNNNDNN-NNNDNNNDNSNNSNNNNKLGIYNVTHNNILKDNKMKAAFKCDNDIYNHSNK 2699

Qy     1081 -------NFNNNEYLSYFQKNVDTIINN---------------CLN---------SLDIS 1109

Db     2700 KSTKQKNNRNKSSYIKLKKYNKSSFINNIIFDNLNKYDKISISCLNLIISIVRLVLDIN 2759

Qy     1110 SMYDDTKEILNNILLSKYKAEKDNVIKKYINE--------DIKNMSLEEIDKTAQSIYEK 1161

Db     2760 SFF----------FLSK---------IIKYIHDLCISTNISEVKFLSFRII---LQMVVNS 2798

Qy     1162 RKVLLTKLLLL-----------FKKNVDTQ--INNETSDLRKDLVMCHICNNNPDDQFHF 1208

Db     2799 RSIKMRRVKRVSSEEEYCDKGSFYSNIEEREWKKEETKEVKEGQYKEDILNDEQNNNTSL 2858

Qy     1209 YAY--SRLEKDIINLIMLRQIWCESENLRLLYQFLV---VEYQNKSANSVLLNVSSNN-- 1261

Db     2859 LSYNSSNFER-YFSLLF-------SPNSTVSYGTLFDLKNEQKGKLENKSNYNIERNNMD 2910

Qy     1262 ----------GDIILLNKKLVQD----------NIKNSMDHNNIH 1286

Db     2911 KATDLYDNYVSDMTKENRTLGHEENEKGTPKYNNNNNNNNNNIN 2955
```

RESULT 18
A23770
asparagine-rich protein - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 09-Jun-2000
C;Accession: A23770
R;Stahl, H.D.; Bianco, A.E.; Crewther, P.E.; Burkot, T.; Coppel, R.L.; Brown, G.V.; And
Nucleic Acids Res. 14, 3089-3102, 1986
A;Title: An asparagine-rich protein from blood stages of Plasmodium falciparum shares d
A;Reference number: A23770; MUID:86176787; PMID:2421257
A;Accession: A23770
A;Molecule type: mRNA
A;Residues: 1-537 <STA>
A;Cross-references: GB:M24328; GB:X03716; NID:g160091; PID:g160092

```
  Query Match             8.7%;  Score 597;  DB 2;  Length 537;
  Best Local Similarity   28.5%;  Pred. No. 3.1e-18;
  Matches  212;  Conservative   92;  Mismatches  193;  Indels  246;  Gaps   39;

Qy      424 NDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNNDSSSMV 483

Db        2 NNNNKNNNNNDDGNINYQNT--NEFKDNKK--NMNFKNQYNNNYKFDENMNNSNTMHSRN 57

Qy      484 NNMINHMINNNIN-NNINVNN-----------NNNNNNNNNNSHNNHLPQPNYAFTDTS 530

Db       58 SNVEEHLRNNSIDMNNSNINNYTNQQTRFSSFMENENENENKNYHTG------------- 104

Qy      531 DFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKL 590

Db      105 ----------GMNNN-----IHFKNKYDN----NNSSMKNTDNNK--------------- 130

Qy      591 TMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNV-----LLGLNKNTHH 645

Db      131 TDTSYN---------MKGTIN--NDNNNMDYLRNINNINEYKGSAKNKFYTNYMNKNNLK 179

Qy      646 DISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMN-SNSI-YKSNYNSNQS 703

Db      180 FTQNNNDNMNINEDNN------NNNNNNNNNNGVFSNYQNNNMNRNNSINIKRNLNNNNN 233

Qy      704 ISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENY 763

Db      234 INN----NMNKMGSQDKNQNS--------NNNFYMNYNYQNRKNSMNNNMNNN------- 274
```

```
Qy        764 NNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDK 823
Db        275 -----MNNNM------NHNM---NNNMNHNMNNNMNHNMNN---NMNHNMN------NNM 311
Qy        824 NYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMK 883
Db        312 N-------------NINSLDSDMSPNYHAHVKMSMMN-YNNN-----ESNTANPN--QMN 350
Qy        884 YGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTF 943
Db        351 F----EQTNNDNMKR---ENNNMNNYGYDDNTVHVNNNTPSTDFFSRAVGYN-------- 395
Qy        944 QNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLST 1003
Db        396 -NNYLNNNN----------NMNSAVNNNS----------------------------SN 415
Qy       1004 NEDIYNKEMEGKHIRLDDQDKYDDND--NNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNN 1061
Db        416 GNNMKNENSENKNVA-------DNNDSLNNNKNNNNNINMNESINNNNTLNN--NNEYNN 466
Qy       1062 VDNNNVDNDDDD---------VDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSL-DISSM 1111
Db        467 -QNNNEDEDDDDWGELGEDKYIDINSIMK----------KKNV--ILNQLEADLNDLSKK 513
Qy       1112 YDDTKEILNNILLSKYKAEKDNV 1134
Db        514 GNDGKN------KKKNKMKKDDL 530
```

RESULT 19
T18427
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18427
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18427
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3724 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1
C;Genetics:
A;Introns: 307/1; 1545/2
A;Note: C0335c

```
  Query Match             8.6%;  Score 591;  DB 2;  Length 3724;
  Best Local Similarity   22.3%;  Pred. No. 4.3e-17;
  Matches  342;  Conservative  244;  Mismatches  543;  Indels  408;  Gaps   82;

Qy         23 DSDIITVKNNCTLYIDEPRYKVDMTKY-IERHEFIVDKVFDDTVDNFTVYENTIKPLIID 81
Db         13 DKDVTNLKQ----YFEGGDIKFENAELKKEREEIRKNKVFIECVQN--IWMNILNKKMDD 66
Qy         82 LYENG-------CVCSCF--------------AYGQTGSGKTYTMLGSQPYGQSDTPGI 119
Db         67 KFSKVEYFKIMLRICKVLIPQFDIKEIIKIVNDEWINDSEGKKYLNFSS----------F 116
Qy        120 FQ--YAAGDIFT-FLNIYDKDN-TKGIF--ISFYEIYCGKLYD--LLQKRKMV-----AA 166
Db        117 FNAFFELSDIWTPTINAYDYAHFLKSLFYRITYIEI---KTKDGRIIKKKPMIMINFKRT 173
Qy        167 LENGKKEVVVKDLKILR---VLTKEELILKMIDGVLLRKIGVNSQN-------DESSRSH 216
Db        174 AEN-KKLSLYNDTNIIEKHVIKEKNKNTLKNLQS--FKRSSMQIQNLKKKHFIKRFSENY 230
Qy        217 AILN---IDLKDINKNTSLGKIAFIDLAGSERGADTVSQN-------------KQTQTD 259
Db        231 KLENSFLYDMDDDGDNSLIPKNLFENLLNNKQHNDYLQRNIILMDVNDINPLEHPDEQNQ 290
Qy        260 GANINRSLLALKECIRAMDSDKNHIP----FRDSE-----LTKVLRDIFVGKSK----SI 306
Db        291 SLNKNKCLTGTNK------KEKYMIPKGTEYQDKEKESSILTINQNDKKYKKKKKKYSEL 344
Qy        307 MIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASS 366
Db        345 QDSNIS-----SNNTLTSKRYTYTCGMDKETKETDEQNKRRNTEVASICNVCNVETNEKQ 399
Qy        367 --------------------IENVVI----------KSNHLLSNNNNNKINRGKINDKI 395
Db        400 RKKTNKKKSVQKDDLVEENALLDNLIIIDGINFDDVTKECKIIDDNNNENIDN---NIYV 456
Qy        396 ERNNILKNKSFDK-PREGFTSTFGKYSSLNDIDKIKKNKKKGLINYK-------STLYND 447
Db        457 ENNKLKDQKSYDLFSSEGKNLILGVNEGEEFNEEVFENIEKELQERKDEKNKNDKTIYNN 516
Qy        448 N---------TINKKHNNNNNNNNDNNDNNNDNNN------------------------- 473
Db        517 NEEQTDDLNRNINKIESINNNNDNNNNINNKKEFNKIRTEHILNKESISKHIKGSPSRDK 576
Qy        474 --------NNNNDSS--------SMVNNMIN---------------HMINNNINNNINV 501
Db        577 EIKKLYTTNKNEDSTFELKKELEIITNNKVNVYEEDIIGSNEDDEYIHVLKENLKE--DA 634
Qy        502 NNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIK 561
Db        635 NEYNNDKENKNNKTKEILKSKNY---r-------------LENEKRTL--------EELK 669
Qy        562 LKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMED-NTPKD 620
Db        670 LRG-----------KNNIFKKDEKYNSLGEVIINEIQINEEN-----KINDIQDGNISKQ 713
Qy        621 ILYESRNVSNMNGNVL-LGLNKNTHHDISTKDENH--------NDNKINNGVINIINNSN 671
Db        714 KIIQSSSRTNDTFNIKDISLNDDLEKEKRKKKSQHFIDNLVKADKNEISEN-INKICDNN 772
Qy        672 VNSINNSNMNSINNSNMNSNSIYKSNYNS--NQSISDVQIRYVNEMDTSNKNN------D 723
Db        773 INNIYDESINNIYDESI--NNIYDESINNIYDESINNIYDENINNIYDENINNIYDENIN 830
Qy        724 NIFFDAIS--CDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNN---SMKLYAY 778
Db        831 NIYDEGINKICDDNIL---------ENKNIKTTNDIVQVEENNESIEKNELMISLNKDIN 881
Qy        779 NSHNLFQPD-----NNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKE 833
Db        882 NTYNMFKENVDIFINKIKRESLLKIDKNIKDNNNDDEYIMDNFYENDFIINHKMEITNKE 941
Qy        834 QKDKNIHGCDNNIIQNRNDFEKKKKTN----------FYNNNNIVIVNNNMGNNNSPRM- 882
Db        942 LDPLEIN-TQNEFIENL-DIKKKKYTNDHFFNDADKMFYEMNK--ILNKDMKKNKEQEFF 997
Qy        883 ----KYGLCGSHTSIDNMKNNEM--KNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNN 936
Db        998 KTDETFGSLQSHKIKKYNKGEEKHDKNNEEEKNILYD------ENQVYSVLYSDHKIEQD 1051
Qy        937 INDDDTFQNDYCHNDNTFTIRRKNN---TNINSNIYQNDDIIYTINSL---NDYMSNTLL 990
Db       1052 IQDIHSIQTNICDENNIEQINEENSKKGVRISGTDMENKNDMENKNDMEKKNDMEKKNDI 1111
Qy        991 HFKEKYTYPTLSTNEDIYNK-EMEGKHIRLDDQDKYDDNDNNN-------VDNNN----- 1037
Db       1112 EKKN-----DMEKKNDMEKKNDMEKK----NDMEKKNDMENENNMENKSDIENENKMNEY 1162
Qy       1038 KNNVDNN---VDNNNVDNN-VDNN--DKNNVDNNNVDN---DDDDVDFHNIKN--FNNNE 1086
Db       1163 KNNMKNENKMIENNTIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNT 1222
Qy       1087 YLSYFQKNVDTIINNCLNSLDISSMYDDTKEILN--NILLSK--YKAEKDNVIKKYINED 1142
Db       1223 KIAY---NMNENKNNI-------KIEDDAKNQINVQNNIQNKVHVKDKKNDLIN---NVD 1269
Qy       1143 IKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNP 1202
```

```
Db       1270 IINDVLKSDDKFENMINSKE--INIKDFKCVDKNNDICLN--LNELNKQSY-----NSNT 1320

Qy       1203 DDQ---FHFYAYSRLEKDIINLIML-RQIWCESENLRLL--------YQFLVVEYQNKSA 1250

Db       1321 TEKCIDIHDFDYVE-NKDVHDKIHEDRKEYCDESKLKYLPADDIKKMRSFIKINKKSKRE 1379

Qy       1251 NSVLLNVSSNNGDIILLNKKLVQDNIKNSMDHN-NIH 1286

Db       1380 N--FLIISYMNEEIYFLIKKLKVEIIEPSIILNKNMH 1414
```

RESULT 20
H71621
serine/threonine-specific protein kinase (EC 2.7.1.-) PFB0150c - malaria parasite (Plasm
C;Species: Plasmodium falciparum
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: H71621
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: H71621
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2485 <GAR>
A;Cross-references: GB:AE001376; GB:AE001362; NID:g3845108; PIDN:AAC71820.1; PID:g384510
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0150c
C;Superfamily: malaria parasite serine/threonine-specific protein kinase PFB0150c; prote
C;Keywords: phosphotransferase
F;2087-2352/Domain: protein kinase homology <KIN>

```
  Query Match             8.6%;  Score 587;  DB 1;  Length 2485;
  Best Local Similarity   20.5%;  Pred. No. 4.1e-17;
  Matches  356;  Conservative  249;  Mismatches  517;  Indels  616;  Gaps   85;

Qy         21 KKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDTVDNFTVYENTIKPLII 80

Db        359 KENINIISNKESC--HKDEK----EEEKYIMYNSNLVEE-----KKQKKMIWNSLNVLPI 407

Qy         81 D-LYENG--------CVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAAGDIFTFL 131

Db        408 DILLKNGHDEINKEICKKKKKSFFSQNDIKSKMLYNNKSYSKSEKV-------------- 453

Qy        132 NIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRVLTKEEL- 190

Db        454 -LYTNNKNSNTFIPIF--FLNKVGDKFKNSENIYDMYNNKKNVYIHDKKIYTNMYSNKLK 510

Qy        191 ----------------ILKMIDGVL-----------------------------LRKIG 204

Db        511 QKHYYSTSNINLLYNNIGKVLDNGLHLSNNMYCRLNSNPPYKSISLINNNVFFYKKRKSN 570

Qy        205 VNSQND----------ESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNK 254

Db        571 SNNNNNNNNISSSSSSSSKKNHVIIN------------KKISSYNIHYKER-KDSFKEN- 616

Qy        255 QTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIA----N 310

Db        617 ------------FLFFKE--KILPSKKDTCVFNERQ-----KDLFEKSNEHIKCVSSFNN 657

Qy        311 ISPTISCCEQT----------LNTLRYSSRVKNF---KNKSTCINEEDDTNTERISILDS 357

Db        658 TSDDISSHSSVNKKEPFFALKNNSIRHIPKENNIIYTSGKS--FNHVQD--KEKTVLLKK 713

Qy        358 KGSEMNASSIENVVIKSN---HLLSNNNNNKINRGKINDKI----ERNNILKNKSFDKPR 410

Db        714 KKEINDKNTFSSCLINHNITTYTLQNGVNKNLNMLGIRDSIYKIDEKNNMLKE------- 766

Qy        411 EGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDN-----TINKKHNNNN------- 458

Db        767 --------CYNGNNDSNNKKKKKKKKL-SFSCDIINDNITPYESDKEKNNSNNIKSMDIF 817

Qy        459 ----------NNNNDNNDNNNDNNNNNNNDSSSMV--NNMINHMINNNINN--------- 497

Db        818 NYVKRKSNLYNNLSSNRDSTVDMHNKYNSEEYINIQRTNKIYELSNKRIRNYKLYSMDEI 877

Qy        498 -NINVNN----NNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLH 552

Db        878 FKVSLKEKKYIDNISNNMERVTYKNEM--INEKISKMDDILYPCDKNKSLNMSCPVI--- 932

Qy        553 KKNLRDNI-KLKNRSSCDNIMNKKKN----------NLHLARHSVGSKLTMFSYDPQKNK 601

Db        933 ---IENNISREENEKNSSVILNKKKNENMFNCVGRLHCHMGK--------------MNNQ 975

Qy        602 DNTFFKSNINKMEDNTPKDILYESRNVSN-MNGNVLLGLNKNTHHDI---STKDENHNDN 657

Db        976 DNIYDQGNIKKNEEEITKHDEYISREEKNKYNSKCIRNFDDYKYEQVLSYHTLDEDKKKN 1035

Qy        658 KIN--------------NGVINII------NNSN-------------------------- 671

Db       1036 DMNNLIDMNNEAIIETVNGVINNIILDRKDNNSRKDMEKEMEKEMEKKMEKEMEKVMEKE 1095

Qy        672 -------------VNSINNSNMNSINNSNMNSNSIYKSN--YNSNQSISDVQIRYVNE-- 714

Db       1096 MEKVMEKEVEKELKNEMNNRMNNRMNNEMKNEINIYKNNEIYVDN----DKELEIVNEEK 1151

Qy        715 ---------------MDTS-NKNN-----DNIFFDAI--SC----DNNMYPNITNNNNNN 747

Db       1152 KLIYPFNYESDVHKNMNMSININNCKDDYNNILKEYVDNSCLAQKEENIFRPLFNLNKKD 1211

Qy        748 N-----NNNNN-----------------NNNNIDVENYNNRD------------------ 767

Db       1212 KVWKRFNIKNNIKTIIHNEEMKRIYQTINKNVFPIYNFNRYENFLINHLTYNFPKNDLFK 1271

Qy        768 -----GTNNSMKLYAYNSH--NLFQPDN---NKN-------TSNIQN------------- 797

Db       1272 LSYKVSMNNIRNLYIANKHINNNYDYMNKLYNQNIYTLKYQVANIDNDHHICKKGGGLDY 1331

Qy        798 INTN-----KNNQDGNVNYSMNFCHYN--------LNDK----NYLIDLNNKEQKDKN-- 838

Db       1332 INMNISKECKNRKD--KTYLNKIFHYKKKKDARFFINDEIGSNDYMYDIKKKYSNDENNY 1389

Qy        839 ----------------------IHG-----CDNNIIQNR--------------NDFEKKK 857

Db       1390 KLNEKMNISMSNDEDMIPTLNSEHGNNFPSCQPNLLEKKSTYIDLNLYDSNSMDDF-TEE 1448

Qy        858 KTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKD--------- 908

Db       1449 KYNFVNNENDLF--------NTKRWKF----NFSKGKNLFNNKFFNVSNEDGVFSFFKNM 1496

Qy        909 NEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNI 968

Db       1497 NLFRELNKSNNSLKLESVKNSNNNCSNN-KGDDNIGNMENMNTTNVTIASDEHISTKGDI 1555

Qy        969 Y-------QNDDIIYTINSLNDYMSNTLLHFKEKYT----------YPTLSTNEDIY--- 1008

Db       1556 HDESFSRDDNDCILLKIEGRSKKYSDITLYNEDKSNLENDNETINEYENVCSNIDVNEWE 1615

Qy       1009 -----------NKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNN 1057

Db       1616 DKVNGTCNSVGDKETEKNNEKNNEKNNEKNNEKNNEKNNEKNNEKNN-EKNNEENNEGNN 1674

Qy       1058 DKNNVDNNNVDNDD-DDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTK 1116

Db       1675 EENNEENNEENNEENNDIEKNDIKDNNSGQ----VKENI--IVMNNTNNMDVDN--DDNN 1726

Qy       1117 EILNNI-------LLSKYKAE------KDNVIKKYINEDIKNMSL--------------E 1149

Db       1727 NNYNNVSTDEGIDIIKNIKSEMNDYIYNDNIMIKINNKSIDLMNIKNQKNEPFLNYTNEK 1786

Qy       1150 EIDKTAQSIY---------------EKRKVLLTKLLLLFKKNV---------DTQINNET 1185

Db       1787 DIHMKSNSSYNVNDKMNLFNNNEKTEKNNTSLNDLLYKRKEELDDEKISEYKDTNLTNNT 1846
```

```
Qy       1186 SD---LRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLR---LLYQ 1239
Db       1847 FEHIAKRINLIL-----NDTIEFFQKHTYLHNGYGNVQVCKKNKRKLEKKKLKKWSCIYK 1901

Qy       1240 FLVVEYQNKSANSVLLNV-SSNNGDII---------LLNKKLVQDNIKNSMDHNNIHK 1287
Db       1902 --INKIVRKGAHGVVFSAWRSENVDFFNHSFFENLNLENKK------KGYIDETNVNE 1951

RESULT 21
T18444
hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18444
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18444
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1650 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1427940; PIDN:CAB11112.2
C;Genetics:
A;Map position: 3
A;Introns: 1597/3; 1625/3
A;Note: C0385c

  Query Match             8.5%;  Score 584;  DB 2;  Length 1650;
  Best Local Similarity   22.4%;  Pred. No. 3.6e-17;
  Matches  263;  Conservative  202;  Mismatches  379;  Indels  328;  Gaps   58;

Qy        328 SSRVKNFKN-------KSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSN 380
Db          2 NEHIKSFKNDNVHYTKKNVIRNYDDHNNIKANSHLKLQSKEYTTHNDE----KQRWITSN 57

Qy        381 NNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINY 440
Db         58 NTNNNIQ--------THANV---KCDDKSVENKTTTLIN-KDINEITNISKQDK----NY 101

Qy        441 KSTLYNDNTIN------------KKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMIN 488
Db        102 --TKYNCLNFNKSDVMRDYVRYSKKEDSSSNNMNGNNNMNGNNNMNGNNNMNG--NNNMN 157

Qy        489 HMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKS 548
Db        158 G--NNNMNGNNNMNGNNNMNDNNNMNGNNIMNGNNI------------MNGN-NNINRN 201

Qy        549 IFLHKKNLRDNIKLKNRSSCDNIMNKKK-NNLHLARHSVGSKLTMFSYDPQKNKDNTFFK 607
Db        202 NIMNGNNIMNG---NNNINRNNIMNRNNITNNHIIHH-ISNKTSLIE-NNKKKEENIFPS 256

Qy        608 SNINKMEDNTPK-----DILYESR-----NVSNMNGNVLLGLNKNTHH----------- 645
Db        257 ANLYKKEMNVKGSSDTFDLLYKRKIDKDDNLSKKKKRPNYLFNKDGEHFVNKENVQNNII 316

Qy        646 --DISTKDENHNDN------KINNGVINIINNSNVNSINNSNMNSI-----NNSNM--NS 690
Db        317 DDDDDDDDNHHDNVVVVYDKVKENEMN--ENKNKKSVKEDGLHNVLVELRNKDNLVVND 374

Qy        691 NSIYKSNYNSNQSISDVQIRYVNEMDTSNKN-NDNIFFDAIS----CDNNMYPNITNNNN 745
Db        375 NIINKSFEKNN-------ILYIKTSDSLNENYNERKIYKEINKEEYSNKNEYVHFKNNDD 427

Qy        746 NNNNNNNNNNNIDVE----------------NYNNRDGTNNSMK-------------LY 776
Db        428 SSIKKKNNSSECLDEQKKKTYKYTIIEQKRYNFNDRDNNNAYIKDDTHKKEKGYYLNMIV 487

Qy        777 AYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLN--NKEQ 834
Db        488 QSEEYKKYGSNNKMDEMEIYNQHTN----DFNIN-------ENLNNKIYFDDYEGYDPEK 536

Qy        835 KDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSID 894
Db        537 KKKKL---DDHIYTQQKEYK--------NNINDILKDHHLNDKETKEKK-----NEIEIE 580

Qy        895 NMKNN-----EMKNNEMKDNEMKDNHI-------------------------KSNNNNSS 924
Db        581 EEKKNKIEIEEEKKNKIEIEEEKKNKIEIEEEKKKKIEIEEEKKKKIEMEEEKNKIDDEK 640

Qy        925 SSSSSNNNIYNNINDDD--------------------------TFQNDY---CHNDNTF 954
Db        641 KNTYANDKIISHIDNVNCNIKIDALLDHIEEKKKTGHKEINLYKEIKNEYQKMLNDENSI 700

Qy        955 TI--RRKNNTN-INSNIYQNDDIIYTINSL---NDYMSNTLLHFKEKYTYPTLSTN--ED 1006
Db        701 MLEHEKKYNTHQVNNNLCDTKDMLQKENKILTNNDKKKTFLLSKSKNITSNVLSSKIPGT 760

Qy       1007 IYNKEMEG------KHIRLDDQDKY-DDNDNNNVDNNNKN--------------NVDNNV 1045
Db        761 LSTKKLNATIKTIKKDVTDNEKKKYVHDHRKDNIIKRNKEFINIYKGKRNYANVEIGSEV 820

Qy       1046 DNNNVDNNVDNN---------DKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNV- 1095
Db        821 CNNKVNVKGDDNKMMVENKQGDDNNMMVENKQGDENNVIVKNIERSESSFMFTHHRKNIT 880

Qy       1096 ----DTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKN--MSLE 1149
Db        881 STSTDTCAKN-KEQIKYPHLYSNKNE----------DKDKKSIFLKNINENIKKNYKDKE 929

Qy       1150 EIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQIN---NETSDLRKDLVMCHICNNNPDDQF 1206
Db        930 KISTLEKKVFVKKDNVITN--NDDEKHTSSKINDDFNITIDKQKGKL-----NNNPVD-- 980

Qy       1207 HFYAYSR--------LEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVS 1258
Db        981 ----LNRKIKNETKILEKDKSHMSKIQN---------NLMKKKTNFPTNNKGISSTSISSS 1028

Qy       1259 SNNG----DIILLNKKLVQDNIKNSMDHNNIH 1286
Db       1029 STKNFKDCGIIEKNKNL--SNLKYTCIRNKMN 1058

RESULT 22
T09079
probable chloroquine resistance protein CG2 (strain 7G8) - malaria parasite (Plasmodium
C;Species: Plasmodium falciparum
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C;Accession: T09079
R;Su, X.Z.; Kirkman, L.A.; Fujioka, H.; Wellems, T.E.
Cell 91, 593-603, 1997
A;Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resistan
A;Reference number: Z16556; MUID:98054002; PMID:9393853
A;Accession: T09079
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2708 <SUX>
A;Cross-references: EMBL:AF030692; NID:g2642513; PIDN:AAC47853.1; PID:g2642514
A;Experimental source: strain 7G8; from Brazil
C;Genetics:
A;Gene: cg2
C;Keywords: toxin resistance

  Query Match             8.5%;  Score 581;  DB 2;  Length 2708;
  Best Local Similarity   22.3%;  Pred. No. 8.1e-17;
  Matches  341;  Conservative  233;  Mismatches  508;  Indels  448;  Gaps   77;

Qy         20 KKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDTVDNFTVYENTIKPLI 79
Db       1219 KQNYLNFIKQKKKCAL---------DLQYYINKRK---------NANYYEYPENYINNYI 1260

Qy         80 IDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPG-------IFQYAAGDIF---- 128
Db       1261 LHLY------------TSFNKKYAMKNEQDMNKDDKNNELMIYLLIKIHSSGSLFFKIP 1307

Qy        129 ----TFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRV 184
```

```
Db     1308 IKGKRYMKINNLDVIKNRELFIYH-KCLVINMKLEKKKKKKKYDNGKKN---DDINTICM 1363

Qy      185 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAIL--NIDLKDINKNTSLGKIAFIDLAG 242
Db     1364 IDQNKLL----------------DNENKIKMNEIIDTNVMNKDICSN----------MVD 1397

Qy      243 SERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGK 302
Db     1398 DEKNEQDTSNNTITSNEG-KINNWVLIKKMYILLND-----------ECTKL-------- 1437

Qy      303 SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCIN-------------------E 343
Db     1438 -------NILDIFSCIIKSCYTFDFSNELYYLENKILCSNFVVRNSHISNIDIVYYSNYN 1490

Qy      344 EDDTNTERISILDSKGSEMNASSIENVVIKSNHLLS-----NNNNNKINRGKINDKIER- 397
Db     1491 KDITCSISLNFLNNEYDNNNNNNIKNELPKSNEMIQKEEILNNNHNTFKQKNRNDMTTPF 1550

Qy      398 ---NNILKNK-----SFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNT 449
Db     1551 STCNDILKNLLFIDISFNNPIDCINNIFIK---------EKNNLYLYLEKKKSIYNLLK 1600

Qy      450 I----NKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMIN-NNINNNINVNNN 504
Db     1601 LICLTYEFHHNFYILINKTNEHLRKTHFLNIMDPSMLFDIQYENLFTINLICKDLYIYNN 1660

Qy      505 NNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKN 564
Db     1661 NNNNNNNNNLCLCILLHPEY----YSQIIIIPYLTCYNNKNVRNDVSFKKK---NYFFKN 1713

Qy      565 RSSCDNIMNKKKNN----LHLAR-HSVG-----SKLT----MFSYDPQK----NKDNTFF 606
Db     1714 FKYDDHLRNKQQNYNGSFLNIQRDHSKGVLERDGKITRDDLLSSFSVQTYTSLNNQNVF- 1772

Qy      607 KSNINKMEDNTPKDI-LYESRNVSNM-NGNVLLGLNKNTHHDISTKDENHNDNKINNGVI 664
Db     1773 ---IDNMDE---EDLFIYLFLNYCNLINFNI-------TNNNVS---QNYND-KISD--L 1813

Qy      665 NIINNSN--------------------VNSINN-----------SNMNSI---NNSNMNS 690
Db     1814 NKTKNKNEERSFWSSSLKKLLTKFDEEIFSTNDLRVFEEYEMFISNIKYILKMKNKIISS 1873

Qy      691 NSIYKSN-------YNSNQSISDVQI------RYVNEMDTSNKNND----NIF------F 727
Db     1874 EVFFFSPYFLPTVLYNLFEFLRTLGVWLTLLGRLKNDYTDINLRNDRNICNVFQYITAKK 1933

Qy      728 DAISCDNNMYPNITNNNNNNNN-------NNNNNNNNIDV----------------ENYN 764
Db     1934 NSKSWDNQPNPEVKNMNTDENNTTTTKKKDDNDNQNDDIYIHLIWNIYNVRTLYTERLN 1993

Qy      765 NRDGTNNSMKLYA---YNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLN 821
Db     1994 NDRKINNAKRKYETFHTNMDDIFNDDN--NCINIINVEDNKEE--------------NIK 2037

Qy      822 DKNYLIDLNNKEQKDKN--IHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNS 879
Db     2038 DLKYKKLKTNEGEKVDNEFIQVTDNNII----EINPKKKTSTQNEEQ---PNINTINENG 2090

Qy      880 PRMKYGLCGSHTSIDNMKNNEMK--------NNEMKDNEMKDNHIKSNNNNSSSSSSNN 931
Db     2091 NMYTYSLINSTLTLNNIHLKRWKYLINTYCFNNYIMFFQTTQNKYLLNRRLIKKAFFLRS 2150

Qy      932 NIYNNINDDDTF-----QNDYCHNDNTFTIRRKNNTNINSNIY-------QNDDIIYTIN 979
Db     2151 LKFDDFNDIKSYYKKVKEINYC--DSNY----KNNKMDTAQQYIHYLKEDQKEKELINFD 2204

Qy      980 SLNDYMSNTL----LHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKY---------- 1025
Db     2205 HIINYMSNNIWVSQIHY--------LTHGISLY---FENVGEEVDEEDAWFYCLQKKKKI 2253

Qy     1026 --DD---NDNNNVDNNNKNNVDN--NVDNN-NVDNNVDNNDKNNVDNNNVDNDDDDVDFH 1077
Db     2254 NGDDNKMNGDNKMNGDNKMNGDNKMNGDNKMNGDNKMNGDNKMNGDDNKINGDDNKINGD 2313

Qy     1078 -NIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTK---------EILNNILLSKY 1127
Db     2314 GNNKNGDDN-------KNCDD-NKNC----DDNKNCDDNKINSDGEVCGEEYSRKLLILY 2361

Qy     1128 KAEKDNVIK-------------KYINEDIKNMSLEEIDK----TAQSIYEKRKV-----L 1165
Db     2362 ---KDNILKENCEINCSLNIIYKYIRIWLLRLKLQDVDVFFCILNEVINDKKKICEFSSL 2418

Qy     1166 LTKLLLLFKKNVDTQINNETSDL-------------RKDLVMCHICNNNPDDQFHFYAYS 1212
Db     2419 VFNSVLSYYEHLDFWLSKNTDVLCSPFIETQFTLREEKDVINDNINNVEKSDLFLKYKKN 2478

Qy     1213 R------LEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDII- 1265
Db     2479 KNKEIKYNDKKFIEII-IPFIIKEKGEEKNLHMDNILKY-NKNDIKIDYNDMNNSMDIIK 2536

Qy     1266 ----------LLNKKLVQDNIKNSMDHNNI 1285
Db     2537 NEDDKEYLSNYINNNI--KNIKNIKNNNDV 2564
```

RESULT 23
T18402
asparagine/aspartate rich protein - malaria parasite (Plasmodium falciparum) (fragments)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18402
R;Barale, J.C.; Candelle, D.; Attal-Bonnefoy, G.; Dehoux, P.; Bonnefoy, S.; Ridley, R.;
Infect. Immun. 65, 3003-3010, 1997
A;Title: Plasmodium falciparum AARP1, a giant protein containing repeated motifs rich i
A;Reference number: Z18929; MUID:97378065; PMID:9234746
A;Accession: T18402
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3844 <BAR>
A;Cross-references: EMBL:Y08926; NID:e1154302; PID:e311435; PIDN:CAA70130.1
C;Genetics:
A;Gene: aarp1

```
  Query Match             8.5%;  Score 581;  DB 2;  Length 3844;
  Best Local Similarity   20.7%;  Pred. No. 1.2e-16;
  Matches  315;  Conservative  234;  Mismatches  474;  Indels  502;  Gaps   63;

Qy       45 DMTKYIERHEF-IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTY 103
Db      898 DMYVEIEIHHIKMADKIFSDIYTNKSTVDSS----------------------------- 928

Qy      104 TMLGSQPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKM 163
Db      929 ------------------KIGHIFYFQFFYDI-----IIRIFNELYYLEYSNIPRKTSY 964

Qy      164 VAALENGKKEVVVKDLKILRVLTKEELILKMIDGV-------------LLRKIGVNSQND 210
Db      965 EKYIERGKYH--VKKI-LIQLLASKDFQYFQLENVFPKQFRHHPSFYELVNKYGYKTHMP 1021

Qy      211 ESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLAL 270
Db     1022 RS-------NVNLIKLNKNSWYLYDAFWPL----------SAFKDFQT----------AN 1054

Qy      271 KECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSR 330
Db     1055 EKCIKEENSS-----------------YIGCSR---------------EEDKEYLNFSFR 1082

Qy      331 -VKN---FKNKSTC----------INEEDDTNTERISILDSKGSEMNASSIENVVIKSNH 376
Db     1083 KVQNLLFYTFKNSCLVSILLIVFIINDEFQRKIEKKKKEELEKEEMKKHMKADVTGEDDS 1142

Qy      377 LLS----NNNNNKINRG-KINDKIERNNILKNKSFDKPREGFTSTFGKYSS---LNDIDK 428
Db     1143 LLNQKGGTTNNNTCDVGVVINEEANRSS------------GMSQNVVTYESGINMNNDDV 1190
```

```
Qy        429 I-KKNKKKGLINYKSTLYNDNTINKKHNNNNNNNND---NNDNNNDNNNNNNNDSSSMVN 484
Db       1191 LGEERTNNDDINYMDDNDDDNDDNDDDNDDNDDDNDDDNNDDNNDDNNDNDNNDDD---- 1246
Qy        485 NMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNN 544
Db       1247 -------NNDDDNNDDDNNDDDNNDDDNNDDDNN-----------DDDNNDDDNNDDDNN 1288
Qy        545 NDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNT 604
Db       1289 DD---------------------DNDDNDEDSNLH------GSSSNMLNIINSRDNNSV 1320
Qy        605 FFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTK----------DENH 654
Db       1321 TNSSNINTARNNNSTYIIVD-QNGDNETLFAEILINRNPNNSINMEIVNVTDNNVGDENN 1379
Qy        655 NDN-----------------KINNGVINIINNSNVNSINNSNMNSINNSNMNS--NSIYK 695
Db       1380 NRNYRHLNADGLSELIRSLISENNIINNVIDESFLQNNNVGGDNPINNDDIRNRGGSIFD 1439
Qy        696 SNYN-SNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYP-NITNNNNN--NNNNN 751
Db       1440 DNISFDNNSSSD----ELNLMDSTNSNNVRI--NLYGRTNGLYDLNFISDSNEIMGNLNT 1493
Qy        752 NNNNNNID-----VENYNNRDGTNNSM----KLYAYNSHNLFQPDNNKNTSNIQNINTNK 802
Db       1494 HNVNNTIDILIKIIKILSVAMGSAHPFIELNSFYYFNAVNFYGTDNDD--------DSEK 1545
Qy        803 NNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFY 862
Db       1546 NEEIGNEENDKKYIDKKYIDKKY---IDNNEEIEKK-HHCDQNKIYDDVDRLSSVSSTSY 1601
Qy        863 NNNNIVIVNNNM---------------GNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMK 907
Db       1602 D-----ILMNHLEESCMYTSYSDETYEKNGNVYNMREGFNVENLSYDKDILNEKKDIVEI 1656
Qy        908 DNEMKDNHIKSNNN----------------NSSSSSSSNNNIYNNINDDDTFQNDYCHND 951
Db       1657 ENE-KEKYIKINDHTVSRNYGCVSLLKDEPKFSSTDGNNNNIINRCNNNNYYYDNMKHVD 1715
Qy        952 -----------------------------------NTFTIRRKNN---------TNINS 966
Db       1716 VKGGEGEGEGEDSEECQIKESYKKMSECNNKENIIFDSISVLRKNNIKRLKNYMCKNKNC 1775
Qy        967 NIYQND----------------------------DIIYTINSLN--------DYMSNTLL 990
Db       1776 YIYYDDNNNKKKKNKKNVENQEKEFYVLNKIFVHNFINCINNINVNEDKCFQKVRSTILN 1835
Qy        991 HFKEKYT--YPTLSTNE-----DIYNKEMEGKHIRLDDQ---------DKYDDNDNNNVD 1034
Db       1836 RLKEMYSGNYDCKNNNSNNEFIELAKKKQEDLLISMKEQQSKFSHFLEEEYSSEENDSLP 1895
Qy       1035 NNNKNN-------------VDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKN 1081
Db       1896 NGGTEDFEDVDFVDDASSYLDSN-SNNNSDGHEDDN-KNRCTDVKTNKQEEEFSISNVNE 1953
Qy       1082 FNNNEYLS-----YFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIK 1136
Db       1954 KKKKKYMKGITICEYEKIKE---KNCICVLCKEGMSK------NNLLAYICFSSHTNLLK 2004
Qy       1137 KYINED----------------------------IKNMSLEEIDKTAQSIYE-----KRK 1163
Db       2005 KIIKKNKFSFPCKHPSIYTCGHFIHTKCLHNTRILKYRKLFSV-KNEPTLYEFTCPLCRS 2063
Qy       1164 VLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFYAYS----------- 1212
Db       2064 IANSFVVYIPKRYISENDKYRIYDCNEEEFFLDNSNKNLSELLKISKYSEERNYGIKRKK 2123
Qy       1213 -RLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLN--- 1268
Db       2124 EREEKKKSKMIMMRGM--ESDNINEMVSDNINEMASDNINEM---VSDNINEMTSDNINE 2178
Qy       1269 ------KKLVQDNIKNSMDHNNIHK 1287
```

```
Db       2179 MASDNINEMVSDNI-NEMTSDNINK 2202
```

RESULT 24
E71604
hypothetical protein PFB0870w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: E71604
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: E71604
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2380 <GAR>
A;Cross-references: GB:AE001421; GB:AE001362; NID:g3845293; PIDN:AAC71963.1; PID:g38452
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0870w

```
  Query Match             8.5%;  Score 579;  DB 2;  Length 2380;
  Best Local Similarity   21.1%;  Pred. No. 8.5e-17;
  Matches  302;  Conservative  208;  Mismatches  414;  Indels  510;  Gaps   67;

Qy        209 NDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLL 268
Db         20 NIESAEEH-------KNINKNIKNKK--FINIDNSNNCNNSNSNNSNSNNNNNNNNNIVR 70
Qy        269 ALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYS 328
Db         71 NNNNFINA-DKKKNVILNEDDDI----------KNKELV----------DESFVNIFFYE 109
Qy        329 SRVKNFKNKSTCINEEDDTNTERISILDSK-GSEMNASSIENVVIKSNHLLSNNNNNKIN 387
Db        110 NYFKNL------FNLNDVSNNKVINIIEQKEGDERNADN----NLKNKNIVRDNINKIKN 159
Qy        388 RGKINDKIERNNILKNKSFDKPREGFTSTFGKY--SSLNDIDKIKKNKKKGLINYKSTLY 445
Db        160 TRNVNEILIYNN-------------------KYIINFLNDTTKCKIEIANFISFYFFFLH 200
Qy        446 NDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMIN-NNINNNINVNNN 504
Db        201 IKDILNK---------------NNDNGLMNKKKSS------LKDICNIKYIYKKIKTSKK 239
Qy        505 NNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIF-------------- 550
Db        240 YISSNDMDTCIRNYL-----YHIDKKNYPIIKKTKCPFLSNTKVLYNKRGYMASCPLTVK 294
Qy        551 ---LHKKNLRDNIKLK---NRSSCDNIMNKKKNNLHLARHSV-------GSKLTMFSYDP 597
Db        295 GKIKHKTNISSKIKLKRERNDSNMFNNMIRKDNNMNVKQEQIHNNDTVNNNMTTNVDGCS 354
Qy        598 QKNKDNTF-------FKSNINKMEDNTPKDILYESRNVSNMN----------------GN 634
Db        355 EPTHDNTFLNIEEEEFKMLKNYLKDVKERKKKYKKGYISTSNFISHGVRLGTTRSRIRGK 414
Qy        635 VLLGLNKNTH-----HDISTKDEN--HNDNKINNGVINIINNS-NVNSINNSNMNSINNS 686
Db        415 CLL-KNKKMHMYDDNEELNKKKKKKMNKDDRIENGIMEDVNDKRKLDCDNKIKFNDIEKE 473
Qy        687 NMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNN 746
Db        474 DLN---------------ICDTE----NVDNNSNNNNNN----------------NNNNNN 499
Qy        747 NNNNNNNNNNNIDVENYNNRDGTNNSM--KLYAYN--------SHNLFQPDNN------- 789
Db        500 NNNNNNNNNNNGYKKKIKNKNKNKNKKKNKLNNYNDNFVSVNGSYDNYSIDNNVINDEIR 559
Qy        790 ---KNTSNIQNINTNKNN---QDGNVNYSMNFCHYNLND-------KNYLIDLNNKEQKD 836
```

```
Db     560 EKKKNNKEVK-IMVDKNNDTEKDGNKKYDTSY-SFNIKNTLSKVFYKNYVKRKGMIKQQH 617

Qy     837 KNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNN--------NMGNNNSPR------- 881

Db     618 NNIHNTHN--IHNTHNMH---NTHNIHNEKVVLLDDTKEKADPMNLGISFSPAGLLIPYH 672

Qy     882 -------MKYGLCGSHTS--------------------------IDNMKNNEMK------ 902

Db     673 LGVSSLLIEKNILNMHTSIAGSSAGSICACCLSVGLSVNKCYFLIENIISNVYKHGCYQK 732

Qy     903 -----NNEMKDNEMKDNHIKSNN---------------------NN--------SSSSSS 928

Db     733 LENILNIELNKYLYEDSYIYLNNRIGNVFVGITQILPYYKKLNINNFYDDNDLISAIIAS 792

Qy     929 SNNNIYNNINDDDTFQNDYCHN------------DNTFT---------------IRRKNN 961

Db     793 CNIPMYLSSNIFVNFRNKKCIDGFFSTKKKDFGCPNTRTERIIKVSPFDSDYVGIGNKNN 852

Qy     962 TNIN------------------------SNIYQNDDIIYTINSLNDYMSNTLLHF----K 993

Db     853 SVISPHLIKYNHILFLFICVKNIFHKYINNLWIEKDYLFLIENLKDILERKIFDYYTFVK 912

Qy     994 EKYTYPTLSTNEDIYNK------EMEGKHIRLDDQDKYDD------NDNNNVDNNNKNNV 1041

Db     913 RYFTF--LRKNETIDDKYEEEEYEDEGEEYEEEDDDEEEDEEEYGHNNDNQDDEGDKNKT 970

Qy    1042 DNNVD------NNNVDNNVDNND--KNNVDNNNVDN----DDDD-------VDFHNI--- 1079

Db     971 TNEKNKKKKNKNNNNNNNIFNNNIFNNNIFNNNMNSCVGVSEKDFISTSIVASFANIKRQ 1030

Qy    1080 -----------------------------KNFNNNEYLS--------------------- 1089

Db    1031 MNEKIEKRKNEKKEKKEKLQRKNMNKCSKNRNRNRYINKDSNIHLMNLIRIKFKNLNYMN 1090

Qy    1090 --------YFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEK------DNVI 1135

Db    1091 MNSFEIELYLKINNDIFLQFNKHNYNVQNFYNFSITLIN--IMSKYYSENFYAYNLEKIV 1148

Qy    1136 KKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMC 1195

Db    1149 YKFL---LNNKNFEYIEKQ----YSSKEDM---------NELDILVN--TYDMKYDKIIE 1190

Qy    1196 HICNNN--PDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFL------------ 1241

Db    1191 FLKNNGYLKIDRY-IYFYPKLKTDII-LFFFKEIFLNDNILKIDRKFLKKNITIMIEVLK 1248

Qy    1242 ---VVEYQNKSANSVL-LNVSSNNGDIILLNKKLVQDNIKNSM-------DHNN 1284

Db    1249 EIFFKEYVKRCITKVIFFPVHMKEHDHV-MNKNYYNNQYVNNSNMFNTRGDHNN 1301
```

RESULT 25
G71609
hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: G71609
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: G71609
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2500 <GAR>
A;Cross-references: GB:AE001408; GB:AE001362; NID:g3845238; PIDN:AAC71919.1; PID:g384524
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0650w

Query Match 8.4%; Score 577.5; DB 2; Length 2500;

Best Local Similarity 24.1%; Pred. No. 1e-16;
Matches 248; Conservative 164; Mismatches 389; Indels 229; Gaps 52;

```
Qy     355 LDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKI----NDKIERNNILKNKSFDKPR 410

Db      33 IDEPGVLSNMNSVSNISTSTNNIGTNTMNFNNSKGFIINPFNENYKKNNIC--TYLDHES 90

Qy     411 EGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINK-------KHNNNNNNNND 463

Db      91 TNINGGVNQYD--NHMDQMNQMNQTNQMNQMNQMNQTNQMNQMNQMNQMNQTNQMNQTNQ 148

Qy     464 NNDNNNDNNNNNNNDSSSMVNNMINHMINN---------NINNNINVNNNNNNNNNNN-- 512

Db     149 MNQTNQMNQTNQMNQTNQMNQMNIQHQRNSVNAPNIYIQNFDQNCDIYYNNNGKSNGNLN 208

Qy     513 --NSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDN 570

Db     209 VQQSDNAHNPL-------IYDISEL--YNREKNEEQKTIF------RD--EYSNRTIIKA 251

Qy     571 IMNKKKNNLHLARHSVGSKLTMFSYDPQKN----------KDNTFFKSNINKME------ 614

Db     252 LINKITNTPMINNSVKNIEDTNSSYNTDENVYNVCSMDEYTTNKYISKNYNENDQVIVQG 311

Qy     615 DNT----PKDILYESRNVS----NMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINI 666

Db     312 NNTVPENDNNEIYKKENLSIFQDSLKDNI---VEYNAYHD------SRHHKPIDEQVAHY 362

Qy     667 INNSNVNSINNS-NMNSINN-----SNMNSNSIYKSNYNSNQSISDVQIRYV-NEMDTSN 719

Db     363 INNYYTNNNNDPYNRNSTNNNGIAENNINVNSAF-NQYKENKQYYDLLNTFTGNIMERKN 421

Qy     720 KNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNID---------VENYNNRD---- 767

Db     422 IMMQNVDYNERINGNSI--NIQGSNNQQMNDQLVDNNNVNMCLMQGPYINNHNMKNFYMG 479

Qy     768 ---GTNNSMKLYAYNSHNLFQPD--------NNKNTSNI--QNINTNK--NNQDGNVNY- 811

Db     480 SNYGGNNNMVHNIMGTNNMVHNNMVHNNMGTNNMGTNNMGNNNIGTNNMGNNNMGNNNIG 539

Qy     812 SMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNN---NIV 868

Db     540 TNNMVHNNMGNNN--IGTNNMVHNNMGNNYMGNNYMGN----------NYMGNNYMGNNY 587

Qy     869 IVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNN---NNSSS 925

Db     588 MVHNNMSTNN-------MGNNYMGNDNMRNNNMGTNNMGTNNMGTNNMGTNNMGTNNMGT 640

Qy     926 SSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDII-YTINSLNDY 984

Db     641 NNMGTNNMGNN-----NMGNNYIGNDN-----MRNNNHII-------DYIINYMVNNMVNN 683

Qy     985 MSNTLLH--FKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNK-NNV 1041

Db     684 MVNTMVNNMVNNMVNYMVNNMVNNMVN-------------------NMGNNMGNHNMINHM 725

Qy    1042 DNN-VDNNNVDNNVDNND--KNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTI 1098

Db     726 GNDRIGNYNMGNNLNSNNYMVNNYSNNTYGNNNNDVN----RNMNYN-YAGYNHMYARCL 780

Qy    1099 INNCLNSL----------DISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSL 1148

Db     781 NNNNMNNTPQYIIPDNKNKISAVHPFTKETNTGIILN--NASQDYTLSRSLGYNL-NFSM 837

Qy    1149 ----EEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLR--KDLVMCHICNN-- 1200

Db     838 IQSENDFNSTPSNIEPVNNQPLNVEPAIFE-----AVNLELFDAESLNDQYVCDENSNTD 892

Qy    1201 ---------NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSAN 1251

Db     893 VIKSKPLVDNPLDDEHIYS-EHLNNRILNDETLPSAQLNVETL-----YGEHEYNEQRMN 946

Qy    1252 SVLLNVSSNN 1261
```

```
Db        947 EQRINEQRMN 956

RESULT 26
B71616
phosphatase (acid phosphatase family) PFB0380c - malaria parasite (Plasmodium falciparum
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71616
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71616
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2010 <GAR>
A;Cross-references: GB:AE001391; GB:AE001362; NID:g3845168; PIDN:AAC71865.1; PID:g384516
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0380c

  Query Match             8.3%;  Score 569.5;  DB 2;  Length 2010;
  Best Local Similarity   21.3%;  Pred. No. 1.8e-16;
  Matches  287;  Conservative  192;  Mismatches  404;  Indels  467;  Gaps   59;

Qy        224 KDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQ---TDGANINRSLLALKECIRAMDSD 280
Db          8 KEVEKKEEKEK-----KGAKDKGKDLFSLNKKRERKKKESQKIDRYL------INSCDSN 56

Qy        281 KNHIP----------------------FRDSELTKVLRDIFVGKSKSIMIANISP---- 313
Db         57 KSNYSCCYLNNECFVKNISICKKCMFSYFEFKNVTKVIYMRHGARTPKKKIKNIWPFKEG 116

Qy        314 ----TISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTE--RISILDSKGSEMNASSI 367
Db        117 KGDLTFLGFQQSIKVGEYLRKYYYTFNK---LNKKYNKRERGLRINNKEKGYIKKNKCDV 173

Qy        368 ENVVIKSNHLLSNNNNNKINRGKINDKI---ERNNILKNKSFDKPREGFTSTFGKYSSLN 424
Db        174 KKCKTLYKNKYNNNNNNNNNYVINEKYNGSNKNDYVKNNTYD--NKGYSYLYDLSTSFN 231

Qy        425 D---------------------------------------------------------- 425
Db        232 ELENRKRKLHKFPYLRDFIYYEKYFLKINKRSNKHQRKVFIKIKRRRRNNILKIWIHQHL 291

Qy        426 IDKIKKNKKKGLINYKSTLYNDNTINKK-----------HNNNNNNNNDNNDNNNDNNN 473
Db        292 INKMKKIKNKNMNNYNKCYIKFSSIRKRGYHKMENIECNNKNNDDDNNDDNNNNNNDDNN 351

Qy        474 NNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNH-----------LPQP 522
Db        352 NNNNDDN----------NNDDNNNDNNNNNDDNNNNNNDDDNNYYYYNYNNDETPFNNK 400

Qy        523 NYAFTDTSDFSSL-------DDMNCHLNNNDKSIF--------LHKKNLRDNIKLKNRSS 567
Db        401 SFNYADMLKYTKYYYKNILKDKKNIYTNNKKKELFFPLMEHLYMYKKKLLIN-KMKEK-- 457

Qy        568 CDNIMNKKKN--------NLHLARHSVGS---KLT---------------MFSYDPQKN 600
Db        458 --NIKKKKKKYDKIIKLINKYLCIKTTNSERCKLTAYGIICGILGISEYIYFFFF----- 510

Qy        601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 660
Db        511 --ILFFKSNYDKTNDNNIDTYTKRKEKKKCLNKRSKCFQNWILNRDITSGQYNCID-KNT 567

Qy        661 NGVINIINNSNVNSINNSNMN-------------SINNSNMNSNSIYKSN-YNSNQSISD 706
Db        568 APVKNYIIGENLCGENGCGKNGCGDILRGDILCGDILRGDNNSIPLFRSNRIFCKQS--- 624

Qy        707 VQIRYVNEMDTSNKNNDNIFFDAI-----SCDNNMYPNITNNNNNNNNNNNNNNNNIDVE 761

Db        625 -KITFCDEL--------YIYFNKILKRLQSLD-DMY----------------KINHEVK 657

Qy        762 NY-NNRDGTNNS-MKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYN 819
Db        658 MFGNDKDVLNNSYKKCYDKNDYGSY-PSYNKYSNDYKS-------------------HYV 697

Qy        820 LNDKNYLIDLNNKEQKDKNIHG--CDNN--IIQNRNDFEKKKK-------TNFYNNNNIV 868
Db        698 I-------------KKMKNVKSVQCSNESIILKERQENEKKKKKKKKMENTFINNNNLM 744

Qy        869 --------IVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNN 920
Db        745 YNINVFFDLIINERGN------------FQFFYNNIKKKRQKNEK-------------- 777

Qy        921 NNSSSSSSNNNIYNNINDDDTFQ-------NDYCHNDNTFTIRRKNNTNINSNIYQNDD 973
Db        778 ------GLEEWNVYN------IFQLYMKYILNEFSKFFKLFKFLNKNVENIDN------- 818

Qy        974 IIYTINSL----NDYMSNTLLHFKEKYTYPTLSTNEDIYNK-EMEGKHIRLDDQDKYDDN 1028
Db        819 ---TFNSITNIYNKYYINMVVHRKDCFEKKQIHSKEHMMKKIHLRDKFIEYEKENEIIDN 875

Qy       1029 DNN-NVDNNNK---NNVDNNVDNNNVDNNVDNND-------------------------K 1059
Db        876 CNNINMDNKKKEINNNYNNMIDNNNIEIDMSNNFIFTYYYIFYLLNYMDTYIQFLFYYLK 935

Qy       1060 NN---------VDNNNVDNDDDDVDFHNIKNFNNN--------EYLSYFQKNVDTII--- 1099
Db        936 NTYILFSVVKVAERNSLMLKTLKTKNHYIKKLRNHIIHNSDVYKILNNYYKDEIFIVYDI 995

Qy       1100 ----NNCLNSLDISSMYDDTK-----EILNNILLSKYKAEKDNVIKKYINEDIKNMSLEE 1150
Db        996 TKWTENCMNTTDI--LYNDVKKNTKIDDLENIDIPIITNDKEEY-------HVNNSIISV 1046

Qy       1151 IDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFYA 1210
Db       1047 LKKHNSSVYKLKKKLKNSIIL--------------KDLKK--LNCNFINKNYIHNTNYDK 1090

Qy       1211 YSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNN--------- 1261
Db       1091 HNKIYQDKIKNWTYHPFHNKKKNVKIIKKF-ISAYDAYIYHGVNLNLNFNRAYEKLSQHP 1149

Qy       1262 -GDIILLNKKLVQDN------IKNSMDHNN 1284
Db       1150 PSSIDLIKKEYGQNNYIINGEIKKYEEQNN 1179

RESULT 27
T09080
probable chloroquine resistance protein CG2 (strain HB3) - malaria parasite (Plasmodium
C;Species: Plasmodium falciparum
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
C;Accession: T09080
R;Su, X.Z.; Kirkman, L.A.; Fujioka, H.; Wellems, T.E.
Cell 91, 593-603, 1997
A;Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resistan
A;Reference number: Z16556; MUID:98054002; PMID:9393853
A;Accession: T09080
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2819 <SUX>
A;Cross-references: EMBL:AF030693; NID:g2642515; PIDN:AAC47854.1; PID:g2642516
A;Experimental source: strain HB3; from Honduras
C;Genetics:
A;Gene: cg2
C;Keywords: toxin resistance

  Query Match             8.3%;  Score 567;  DB 2;  Length 2819;
  Best Local Similarity   19.8%;  Pred. No. 3.2e-16;
  Matches  345;  Conservative  244;  Mismatches  490;  Indels  666;  Gaps   78;

Qy         20 KKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDTVDNFTVYENTIKPLI 79
```

```
Db   1226 KQNYLNFIKQKKKCAL---------DLQYYINKRK---------NANYYEYPENYINNYI 1267

Qy     80 IDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPG-------IFQYAAGDIF---- 128
Db   1268 LHLY-------------TSFNKKYAMKNEQDMNKDDKNNELMIYLLIKIHSSGSLFFKIP 1314

Qy    129 ----TFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRV 184
Db   1315 IKGKRYMKINNLDVIKNRELFIYH-KCLVINMKLEKKKKKKKYDNGKKN---DDINTICM 1370

Qy    185 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAIL--NIDLKDINKNTSLGKIAFIDLAG 242
Db   1371 IDQNKLL----------------DNENKIKMNEIIDTNVMNKDICSN---------MVD 1404

Qy    243 SERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGK 302
Db   1405 DEKNEQDTSNNTITSNEG-KINNWVLIKKMYILLND-----------ECTKL-------- 1444

Qy    303 SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCIN------------------E 343
Db   1445 -------NILDIFSCIIKSCYTFDFSNELYYLENKILCSNFVVRNSHISNIDIVYYSNYN 1497

Qy    344 EDDTNTERISILDSKGSEMNASSIENVVIKSNHLLS-----NNNNNKINRGKINDKIER- 397
Db   1498 KDITCSISLNFLNNEYDNNNNNNIKNELPKSNEMIQKEEILNNNHNTFKQKNRNDMTTPF 1557

Qy    398 ---NNILKNK-----SFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNT 449
Db   1558 STCNDILKNLLFIDISFNNPIDCINNIFIK----------EKNNLYLYLEKKKSIYNLLK 1607

Qy    450 I----NKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMIN-NNINNNINVNNN 504
Db   1608 LICLTYEFHHNFYILINKTNEHLRKTHFLNIMDPSMLFDIQYENLFTINLICKDLYIYNN 1667

Qy    505 NNNNNNN---------------------------------------------NNSHNNHL-- 519
Db   1668 NNNNNNNLCLCILLHPEYYSQIIIIPYLTCYNNKNVRNDVSFKKKNYFFKNFKYDDHLRN 1727

Qy    520 PQPNY--AFTD----------------TSD----------FSSLDDMNCHLNNNDKS--- 548
Db   1728 KQQNYNGSFLNIQRDHSKGVLERDGKITRDDLLSSFSVQTYTSLNNQNVFIDNMDEEDLF 1787

Qy    549 --IFLHKKNLRDNIKLKNRSSCDNI------MNKKKNNLHLARHSVGSKLTMFSYDPQKN 600
Db   1788 IYLFLNYCNL-INFNITNNNVSQNYNDKISDLNKTKN---------------------KN 1825

Qy    601 KDNTFFKSNINKMEDNTPKDIL----------YES-----RNVSNMNGNV---------- 635
Db   1826 EERSFWSSSLKKLLTKFDEEIFSTNDLRVFEEYEMFISNIKYILKMKNKIISSEVFFFSP 1885

Qy    636 ----------------------LLGLNKNTHHDISTKDENHN----------------DN 657
Db   1886 YFLPTVLYNVFEFLRTLGVWLTLLGRLKNDYTDINLRNDRNICNVFQYITAKKNSKSWDN 1945

Qy    658 KINNGVINIINNSNVNSINNSNMNSINNSNMNSN--------------SIYKSNYNSNQSI 704
Db   1946 QPNPEVKNM--NTDENNTTTTKKKDDNDNQNDDIYIHLIWNIYNVRTLYTERLNNDRKI 2003

Qy    705 SDVQIRYVNEMDTSNKNNDNIFFDAISCDN---------------------------N 735
Db   2004 SNAKRKY----ETFHTNMDDIFNDDNNCINIINVEDNKEENIKDLKYKKLKTNEVEKVDN 2059

Qy    736 MYPNITNNN-------------NNNNNNNNNNNNIDVENYNNRDGT----NNSMK---- 774
Db   2060 EFIQVTDNNIIEINPKKKTSTQNEEQHNINTINENGNMYTYSLINSTLTLNNIHMKRWKY 2119

Qy    775 ---LYAYNSHNL-FQPDNNKNTSN---------------------------IQNINTNKN 803
Db   2120 LINTYCFNNYIMFFQTTQNKYLLNRRLIKKAFFLRSLKFDDFNDIKSYYKKVKEINYCDN 2179

Qy    804 NQDGN-VNYSMNFCHYNLNDK------------NYLID--------------------- 828
Db   2180 NHKNNKMDTAQQYIHYLKEDQKEKELINFDHIINYMSDNIWVSQIHYLTHGISLYFENVG 2239

Qy    829 --LNNKE------QKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSP 880
Db   2240 EEVDEEDAWFYCLQKKKKINGDDNKM----NGDNKMNGDNKMNGDNKMNGDNKMNGDNKM 2295

Qy    881 RMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSNNNIYNNINDD 940
Db   2296 NGDNKMNG-----DNKMNGDNKMN--GDNKMNGDNKMNGDNKMNGDNKMNGD--NKMNGD 2346

Qy    941 DTFQNDYCHN-DNTFTIRRKNN--TNINSNIYQNDDIIYTINSLN-DYMSNTLLHFKEKY 996
Db   2347 NKMNGDNKMNGDNKMNGDNKMNGDNKMNGDNKMNGD-----NKMNGDNKMN--------- 2392

Qy    997 TYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDN 1056
Db   2393 -----GDNKMNGDNKMNGDNKMNGDDNKINGDDNKNCDDN--KNCD---DNKNCDDNKNC 2442

Qy   1057 NDKNNVDNNNVDNDDD--------------------------------------- 1072
Db   2443 DDNKNCDDNKINCDGEVCGEEYSRKLLILYKDNILKENCEINCSLNIIYKYIRIWLLRLK 2502

Qy   1073 --DVD-FHNIKN-----------FNN---NEYLSYFQ-------KNVDT----------- 1097
Db   2503 LQDVDVFFCILNEVINDKKKICEFSSLVFNSVLSYYEHLDFWLSKNTDVLCSPFIETQFT 2562

Qy   1098 ------IINNCLNSLDISSM-------------YDDTK--EILNNILLSKYKAEK----D 1132
Db   2563 LREEKDVINDNINNVEKSDLFLKYKKNKNKEIKYNDKKFIEIIIPFIIKEKGEEKNLHMD 2622

Qy   1133 NVIKKYINEDIK------NMSLE----EIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQIN 1182
Db   2623 NIL-KYNKNDIKIDYNDMNNSMDIIKNEDDKEYLSNYINNNI---KNIKNIKNNNDVQNS 2678

Qy   1183 NETSDLRK-DLVM-----CHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRL 1236
Db   2679 NMNSNYKNIDINMLNALNCSVIN-----------YNNIK---INYKIWRYMPC------- 2717

Qy   1237 LYQFLVVEYQ-------NKSANSVLLN--VSS----NNGDIILLNKKLVQDNIKNSMDHN 1283
Db   2718 --PFKIISYNKNKLDILKKNNDSIILSFQISSLPVKNLKDVYINNEPPIKFLVTNHIDLN 2775

Qy   1284 NIHKK 1288
Db   2776 VAHER 2780
```

RESULT 28
B71612
hypothetical protein PFB0555c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71612
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71612
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3973 <GAR>
A;Cross-references: GB:AE001402; GB:AE001362; NID:g3845213; PIDN:AAC71900.1; PID:g38452
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0555c
C;Superfamily: Plasmodium falciparum hypothetical protein PFB0555c

```
  Query Match             8.2%;  Score 563.5;  DB 2;  Length 3973;
  Best Local Similarity   20.6%;  Pred. No. 6.5e-16;
  Matches  352;  Conservative  246;  Mismatches  541;  Indels  573;  Gaps   82;

Qy         17 LEKKKKDSDIITVKNNCTLYI----DEPRYKV-------DMTK------YIERHEFIVD- 58
```

```
Db      778 VEKLAQEEGNITKKNHKNIHMVILEDTLYYKIIYILRKKNINKDTIYITYIEYYYILLDY 837

Qy       59 --KVFDDTVDNFTVYENTIKPLIIDLY-ENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD 115

Db      838 FCKYFYNNMDNYFKYNYMKKSTVRKKYVKNKNINDTKGHNKNNNN------NNNIYGDDD 891

Qy      116 TPGIFQYAAGDIF------TFLNIYDKDNTKGIFISFYEI---YCGKLYDLLQKRKMVAA 166

Db      892 DNNIYCHDDDDIYCHDDDDIYCHDDDDDN---IFIFFEKIVIFFC----NILQINKCFHI 944

Qy      167 LENGKKEVVVK-------------------DLKILRVLTKEELILKMIDGVLLRKIGVNS 207

Db      945 LIEIKLNYYFKIPYDIKNTFINLFTYLFFFSLKVQEFSRFQIMIVRCL-SFLLKKKNINK 1003

Qy      208 QNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDG-ANINRS 266

Db     1004 LNAYIFQLFSYLENDQININEK---GMVHRKSSKYHRNNQEEYSHNNKTNDNSVSNLYRD 1060

Qy      267 LLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLR 326

Db     1061 I--------ENEYDENHLERR--------------KDRNVFSSNMNDDKK--YNNLSDFK 1096

Qy      327 YSSRVKNFKNKSTCINEEDDTNTERISIL--DSKGSEMNASSIENVVIKS---------- 374

Db     1097 YTKENMDIKE-----NFRIDISFLKIFFLLNDVRQINLNESNGRKDKLESKAKRRIQKLD 1151

Qy      375 --NHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKN 432

Db     1152 VHRYTYNENDNNKYNDG--NTFLSSQDEEKSKSFDS------------SDSCSVD--EKE 1195

Qy      433 KKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNN---NNDSSSMV------ 483

Db     1196 SSKGLYG------NDFVNSSDHNNNSSNNSSNNSSNNSSNNSSSGRNNSSDEVVVDPYDY 1249

Qy      484 NN------------MINHMINN--NINNNINVNN-------------------NNNNNNN 510

Db     1250 NNYYECKDSNKFGVVVNYFYAHLPNFEKSYNINYVVEDISFDDIFLISIMDLWETNNNNN 1309

Qy      511 NNNSHNNHL-----PQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKL--- 562

Db     1310 LLNLINDLLKIYEEEKKKKIYICTSLLLKIFKRIIKKKSNSYFLFNIYKAFENDIKLILD 1369

Qy      563 --------------KNRSSCDNIMNKKKN---------------------NLHLARH--- 584

Db     1370 SINILIKKWVVWTFKN---CDNIFNREKNINIKKLVKLFFISFYKYLKNYFLQIYYHFFY 1426

Qy      585 -----------------SVGSK------LTMFSYDPQKNKDNTFFKSNINK--------- 612

Db     1427 NNQIYNRKNYNFDNFFFSIFSKYINKIFVEIYSSSSSSTSSNSSFVFNVSKFYMMKMCIS 1486

Qy      613 -----------MEDNTPKDILYESRNVSNMNGNVLLG--LNKNTHHDISTKDENHNDNKI 659

Db     1487 IINNMIGVVKYINLERVKQVFYEH--------NIMMDVHMKSHLHHDIDV--YYGHDNSY 1536

Qy      660 NN-------------GVINIINNSNVNSINN-----SNMNSIN-------NSNMNSNSI 693

Db     1537 NNIYQKIIKSYRGEEKDTLDVINTESVHQNRNEDDIDGSINSLDVFNEIMRNIIINHNSL 1596

Qy      694 YKSNYNSNQSISDVQIRYVNEMDTSNK--NND---NIFFDAISCDNNMYPNITNNNN--- 745

Db     1597 IKDHNDMCTKKKRINIFQISSPATSEQLMNNHYTMNYLTDVMLLQKDYIYNIDNNMNEHK 1656

Qy      746 ---------NNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNT---- 792

Db     1657 QNVFNKPFDNNNNNNNNNNFMLNYFNYIPENNNNN----YRMDIKKRYPPESYDNNYYMF 1712

Qy      793 SNIQNINTNKNNQDGNVNYSMNFCHYNLND--------------KNYLIDLNNKEQKDKN 838

Db     1713 NNIKNEEENILLQNNSMSSSIYIDKKLMKDTKEMEPLFNKTKDMKNY-----NEEQKNNE 1767

Qy      839 IHGCD-NNIIQNRNDFEKKKKTNFYNNN--NIVIVNNNM--------------------- 874

Db     1768 LISYPYNNMLQNNIIF---VKFFLYTQNLLQIIFQNNYIFFLSDFLFINYKKKEYIEEKK 1824

Qy      875 -GNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNNSSSSSSSNN-- 931

Db     1825 NGNQNVINIK----DEDKHITNIKDGDKHITNIKDGDKNITNIKDDDKNITNMKKKNNKN 1880

Qy      932 ----------------NIYNNINDDDTF--QNDY---C---HNDNT-FTIRRKNN----- 961

Db     1881 YLTILMYNSQECSFYYSIFNTLINDYNFLYYKDYKSSCFLYESLNTFFKINNFNNIYFLC 1940

Qy      962 --------------------------TNINSN-------------------------- 967

Db     1941 KYSSGYLPLERIIKLFMDVIFGHFIKFININENINDYELLEVLEYNGNKCYELLRFLFFF 2000

Qy      968 IYQNDDIIYTINSLNDYMSNTLLHFKEKYT-----YPTLSTNEDIYNKEMEGKHIRL--- 1019

Db     2001 IKQND--LITIN-IYKYIFDIIMCILEQYIAHVNYYIYKGKAWDVFFNKLKILNLSLHFV 2057

Qy     1020 ---------DD---QDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNV 1067

Db     2058 NSIYFNIFCDDINAEIKRENDNNKSNDNNNKNNHNN---NNNKNNHNNNNNKNNEKTQNV 2114

Qy     1068 DNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKY 1127

Db     2115 D---------------EQRRSRLWNIVECLFYNVINKLHVNSI-----NCLKKNKLGSY 2153

Qy     1128 KAE----KDNVIKKYI---NEDIK----------NMSLEEIDKTAQSIY-EKRKVLLTKL 1169

Db     2154 KCDEEFPKELNCKRYFLNYNKDFKKEIYYYLYNLNIASEIFELIIKAIYINETKIYPLII 2213

Qy     1170 LLLFKKNVD------------------TQINNETSDLRKDLVMCHICNNNPDDQFHFYAY 1211

Db     2214 NICYDRNISNIFFNIDYDNLNSILEKYTYLHKKKKDHIKNLKYL-LCKNK---SIHMHKY 2269

Qy     1212 -SRLEKD-IIN--LIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLN----VSSNNGD 1263

Db     2270 ISYIDDDHLINNMLHLLRR-----KNIYYKYVLNINEYNNFLDNHKCKRKRKFINYNNIQ 2324

Qy     1264 IILLNKKLVQDNIKNSMDH-------NNIHKK 1288

Db     2325 SSYNNNYNIYNNTNNFYEYHDYIAIKNILHKK 2356
```

RESULT 29
A42640
kinesin-related protein KIP1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YBL0504; protein YBL0521; protein YBL063w
C;Species: Saccharomyces cerevisiae
C;Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C;Accession: A42640; S39826; S45798; S37327
R;Roof, D.M.; Meluh, P.B.; Rose, M.D.
J. Cell Biol. 118, 95-108, 1992
A;Title: Kinesin-related proteins required for assembly of the mitotic spindle.
A;Reference number: A42640; MUID:92317166; PMID:1618910
A;Accession: A42640
A;Molecule type: DNA
A;Residues: 1-1111 <ROO>
A;Cross-references: EMBL:Z11962; NID:g3850; PIDN:CAA78019.1; PID:g3851
A;Note: sequence extracted from NCBI backbone (NCBIN:107715, NCBIP:107718)
R;Scherens, B.; el Bakkoury, M.; Vierendeels, F.; Dubois, E.; Messenguy, F.
Yeast 9, 1355-1371, 1993
A;Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of y
A;Reference number: S39824; MUID:94205266; PMID:8154187
A;Accession: S39826
A;Molecule type: DNA
A;Residues: 1-1111 <SCH>
A;Cross-references: EMBL:Z23261; NID:g313733; PIDN:CAA80785.1; PID:g313736
R;Dubois, E.; El Bakkoury, M.; Glansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45782
A;Accession: S45798
A;Molecule type: DNA
A;Residues: 1-1111 <DUB>
A;Cross-references: EMBL:Z35824; NID:g536098; PIDN:CAA84883.1; PID:g536099; GSPDB:GN000

```
C;Genetics:
A;Gene: SGD:KIP1; MIPS:YBL063w
A;Cross-references: SGD:S0000159; MIPS:YBL063w
A;Map position: 2L
C;Superfamily: kinesin-related protein KIP1; kinesin motor domain homology
C;Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F;53-416/Domain: kinesin motor domain homology <KMOT>
F;141-148/Region: nucleotide-binding motif A (P-loop)
F;147/Binding site: ATP (Lys) #status predicted

  Query Match              8.2%;  Score 561.5;  DB 1;  Length 1111;
  Best Local Similarity   22.9%;  Pred. No. 2e-16;
  Matches  255;  Conservative  204;  Mismatches  393;  Indels  263;  Gaps   48;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITV---KNNCTLYIDEPRYKVDMTKYIERHEFIVD 58
              :| | | || |  :: | :|| | :|:   :    :  :      :|  : ::| |
Db         50 DSNIHVYVRCRSRNKREIEEKSSVVISTLGPQGKEIILSNGSHQSYSSSK--KTYQF--D 105

Qy         59 KVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLG---------S 108
              :||    |  ||:  | |   | ::  :|  |: ||||||:|||||| |         :
Db        106 QVFGAESDQETVFNATAKNYIKEML-HGYNCTIFAYGQTGTGKTYTMSGDINILGDVQST 164

Qy        109 QPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQ--------- 159
                 :  ||       |:|  |:  :|: :  : ||| |:|   | |||
Db        165 DNLLLGEHAGIIPRVLVDLFKELSSLNKEYS--VKISFLELYNENLKDLLSDSEDDDPAV 222

Qy        160 ---KRKM-VAALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRS 215
                 ||:: :  .  |      ::|| :: :  :   :: | |  ||:      || ||||
Db        223 NDPKRQIRIFDNNNNSSIMVKGMQEIFINSAHEGLNLLMQGSLKRKVAATKCNDLSSRS 282

Qy        216 HAILNI-------DLKD--INKN-TSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINR 265
              | :   |      |||    |||   :||:  :|||||    : ::|: |   |   ||:
Db        283 HTVFTITTNIVEQDSKDHGQNKNFVKIGKLNLVDLAGSENINRSGAENKRAQEAGL-INK 341

Qy        266 SLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTL 325
              ||| |    | |:    ||||:|:|:|||::|:|    | :|: :|| |||      |:| :||
Db        342 SLLTLGRVINALVDHSNHIPYRESKLTRLLQDSLGGMTKTCIIATISPAKISMEETASTL 401

Qy        326 RYSSRVKNFKN---------KSTCINE------------EDDTNTERISILDSKGSEMNA 364
               |::| |: ||          | ||: :            ::  | : | |         :::
Db        402 EYATRAKSIKNTPQVNQSLSKDTCLKDYIQEIEKLRNDLKNSRNKQGIFITQ---DQLDL 458

Qy        365 SSIENVVIKSNHLLSNNNNNKINRGKIN--DKIERNNILKN---------KSFDKPREGF 413
                :::|    :|  :|    :| :  | |  :::: ||:|::          ::|:    |
Db        459 YESNSILIDEQNLKIHNLREQIKKFKENYLNQLDINNLLQSEKEKLIAIIQNFNVDFSNF 518

Qy        414 TSTFGK--YSSLNDIDKIKKNKKKGLINYKSTLYNDN------------TINKKHNNNN 458
              |    |  :::|  :::: : :     |  |  ||            |:|    |
Db        519 YSEIQKIHHTNLELMNEVIQQRDFSLENSQKQ-YNTNQNMQLKISQQVLQTLNTLQGSLN 577

Qy        459 NNNNDNND---------NNNDNNNNNNNDSS-SMVNNMINHMINNNINNNI-NVNNNNNN 507
              | |:  ::             | |  :  :||:   : |:  ::: | :|  : ::: :
Db        578 NYNSKCSEVIKGVTEELTRNVNTHKAKHDSTLKSLLNITTNLLMNQMNELVRSISTSLEI 637

Qy        508 NNNNNNSHN----NHLPQPNYAF--------------TDTSDFSSLDDM--NCHLNNNDK 547
                ::: ||       |  : |  : |                :|  :|::::  ||  | |
Db        638 FQSDSTSHYRKDLNEIYQSHQQFLKNLQNDIKSCLDSIGSSILTSINEISQNCTTNLNSM 697

Qy        548 SIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNL-----------------------HLAR- 583
              :: :  :     :  :|:|  |   |  | ||:|                        |::|
Db        698 NVLIENQQSGSSKLIKEQ---DLEIKKLKNDLINERRISNQFNQQLAEMKRYFQDHVSRT 754

Qy        584 ----HSVGSKLTMFSYDPQKNKDNTFFKSN----------INKMEDNTPKDILYESRN-- 627
                     :|  |       |        |            :||:  :: :  : : |
Db        755 RSEFHDELNKCIDNLKDKQSKLDQDIWQKTASIFNETDIVVNKIHSDSIASLAHNAENTL 814

Qy        628 --VSNMN---GNVLLGLNKNTHHDISTK----DENHNDNKINNGVINIINNSNVNSINNS 678
                ||  |     | |: |:: : |||:|     |      |||:  :     :  |  : :|
Db        815 KTVSQNNESFTNDLISLSRGMNMDISSKLRSLPINEFLNKISQTICETCGDDNTIA-SNP 873

Qy        679 NMNSINN-SNMNSNSIYKSNYNSNQSISDVQIRYVNEMDT-SNKNNDNIFFDAISCDNNM 736
               : ||    |:  :  | :|          | ::|      ::::| ||:|| |:
Db        874 VLTSIKKFQNIICSDIALTNEKIMSLIDEIQ----SQIETISNENNINL----------- 918

Qy        737 YPNITNNNNNNNNN----NNNNNNIDVENYNNRDGTNNSMKLYA--------YNSHNLF 784
               | | |:. |      : : | : :    :  :  : || :         : :|::
Db        919 ---IAINENFNSLCNFILTDYDENIMQISKTQDEVLSEHCEKLQSLKILGMDIFTAHSIE 975

Qy        785 QP---DNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKD-KNIH 840
              :|       | |: :           ::|  |  |         | || : |  |
Db        976 KPLHEHTRPEASVIKALPL--------LDYPKQFQIYR--------DAENKSKDDTSNSR 1019

Qy        841 GCDNNIIQNRN----DFEKK----------------KKTNFYNNNNIVIVNNNMG----- 875
                | |: ||   | |              | |  :|   : |  |
Db       1020 TCIPNLSTNENFPLSQFSPKTPVPVPDQPLPKVLIPKSINSAKSNRSKTLPNTEGTGRES 1079

Qy        876 NNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNE 910
              ||  |         |:  :|  | :||::  |:
Db       1080 QNNLKR-------RFTTEPILKGEETENNDILQNK 1107


RESULT 30
A45597
DNA-directed RNA polymerase (EC 2.7.7.6) III largest chain - malaria parasite (Plasmodi
C;Species: Plasmodium falciparum
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jun-2000
C;Accession: A45597
R;Li, W.B.; Bzik, D.J.; Tanaka, M.; Gu, H.M.; Fox, B.A.; Inselburg, J.
Mol. Biochem. Parasitol. 46, 229-239, 1991
A;Title: Characterization of the gene encoding the largest subunit of Plasmodium falcip
A;Reference number: A45597; MUID:92018020; PMID:1656254
A;Accession: A45597
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2339 <LI1>
A;Note: sequence extracted from NCBI backbone (NCBIN:61099, NCBIP:61101)
C;Superfamily: Plasmodium DNA-directed RNA polymerase II large chain
C;Keywords: nucleotidyltransferase; transcription

  Query Match              8.2%;  Score 560.5;  DB 2;  Length 2339;
  Best Local Similarity   20.3%;  Pred. No. 5e-16;
  Matches  356;  Conservative  234;  Mismatches  520;  Indels  645;  Gaps   73;

Qy         31 NNCTLYIDEPRYKVDMTKYIERHEFIVDKVF-----DDTVDNFTVYE-------NTIKPL 78
              :: :|||: |   :    :  | : :: :       ::|: || : |         :| |
Db        684 SDASLYIELPIPAILKPKELWTGKQLISVLIKPNKKENTIINFEIQEREYSNKFGDLKHL 743

Qy         79 IID-----LYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY-----AAGDIF 128
               ::      |:: :|     | |         :|||  |          |:| |       ::
Db        744 CLNDSYVCFYKSELIC-----GSLGK----KVLGSSKY------GLFYYLIHHNSSHIAL 788

Qy        129 TFLNIYDK-----DNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVV---KDLK 180
                :|   |       : ||: |    ::    :      :|  ::    |    |:::  | ::
Db        789 KIMNRLSKLTSRYFSNKGMTIGIDDVRPSQTLTEKKKDLLLKGYEKVNNEIILYNEKKMQ 848

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
              |    | |:         : ::| | :||    : |:  | :       |  |  | ||::
Db        849 IQPGCTLEETL-----EIKVKSILDDLRNDAGKTCNQYLHYLNKPLIMFNSGAKGALINI 903

Qy        241 A------GSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSE---- 290
              |       | :  ||          |||:|                    |  :||
Db        904 AQMIACVGQQNVAGQRIQN-------GFINRTL------------PHFHFHCKDSESRGF 944

Qy        291 --------------------------------------LTKVLRDIFVGKSKSIMIA 309
                                                    | | | |: :    |:
Db        945 VQNSFYTGLSPTEFFFHTMSGREGLVDTAVKTAETGYMQRRLMKALEDLSIHYDYSVR-- 1002

Qy        310 NISPTISCCEQTLNTLRYSSRVKN---FKNKSTCINEEDDTNTERISILDS--------- 357
                   || :| :  : |  |  |   | :| :: |    :| |
Db       1003 ------SCDKQIVQFI-YGDDALNPSYIDNNNTYLDQFDKVFDHIVSISSSHLLLSYKNK 1055

Qy        358 -----------KGSEMNASSIENVVIKSNHLLSNNNNNKINRGKI-NDKIERNNILKNKS 405
                        | | ||      |  :|  ::| |||  || |  |:     ||| | |
Db       1056 IPYLPHVQHQNKTSNMNN-------IYNN--MNNINNNDSNRSIIYNNDSNMNNINNNDS 1106

Qy        406 FDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLY-NDNTINKKHNNNNNNNNDN 464
```

```
Db     1107 ------NMNSIHNNNSNMNNIHNNDSN--------RSIIHNNDSNMNSIHNNDSNMNSIH 1152

Qy      465 NDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNSHN-------- 516

Db     1153 NNNSNMNNIHNNDSNRSIIHN------NDSNMNSIHNNDSNNNNNYKDCTHNPYICNESL 1206

Qy      517 ------NHLPQPNYA-----------FTDTSDFSSLDDMNCHLNNNDKSIFLHKKNL-- 556

Db     1207 IIRNIMNRLIYQNIAQEDLFIPLEHDEFLVNKIMESYTDQECNYEDIIRSLDLN-KNVSY 1265

Qy      557 --RDNIKLKNRSSCD----NIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNI 610

Db     1266 IHNDQGKHLSLQMCAEEHITINNTNNDNTYV------EQIEMKELSKNKTKEKQSFKGTI 1319

Qy      611 NKMEDNTPKDI-------------------------------LYESRNVSNMNGNVLLGL 639

Db     1320 RDMHEDSEEQMNKFITKKAKFFIEKKKGKMHECNDDIEYNNTQYDNIQYNNISCNYIKSQ 1379

Qy      640 N-KNTHHDISTKDENHNDNKINNGVI----------NIINN-SNVNSINN---SNMNSIN 684

Db     1380 NLENTHHQV-----NNDLSFIKNNVILPPKEYHSIFHFVNDYRNVVEIKNLMDKKKIFLN 1434

Qy      685 NSNMNSNSIYKSNYNS-----NQSISDVQIRYVNEMDTSNK------NNDNIFF---DAI 730

Db     1435 NSEKN---VVQSKYNRMSKNLKKKIEIINNIYRNEKKKLNRWKTKMDNDDNYWSSDDDSI 1491

Qy      731 ----------------------SCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRD 767

Db     1492 IAKKIIKIKNKEKRKYHPKEEKENFDRNNYKMITDNNNNDNNNNNNDNNNNDNNNNNNNS 1551

Qy      768 GTNNSMKLYAYNSH-------------NLFQPDNN----KNTSNIQNINTNKNNQDGNVN 810

Db     1552 NNNN----YYYNLHDDVNNLGVTNYNTNIYPNDCNGIYEKETNN-NELTTNSNMCDKNND 1606

Qy      811 YSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDN--NIIQNRND------FEKKKKTNFY 862

Db     1607 FSDEFFN-NINENDLLYDNKYYRQIFKNVIGFVSVFEYVESYKQHYILFPYEIIKWTSFL 1665

Qy      863 NNNNIVIVNNNM-------------GNNNSPRMK-----------------YGLCGSHTS 892

Db     1666 LEYLTEIIPTNIFLHTKLSKKEKPTHQKNTGKMKIYIEEIKKWLFIKAINIYKYFSFKKS 1725

Qy      893 IDNMKNNEMKNNEMKDNEMKDNH------------------------------------ 915

Db     1726 IELIKKKDYFNYIIKNYDISHRYIIHDYSFINLKQLYLFIFFNIYKYFKYISTPGDAVGS 1785

Qy      916 ----------------------------------IKSNNNNSSSSSSNNNIYNNINDDD 941

Db     1786 ISAQSIGEPGTQMTLKTFHFAGVASMNVTLGVPRIKEIINASNSIQTPILNIPLEVNDNY 1845

Qy      942 TFQ-----------NDYC-HNDNTFTIRR-----KNNTN------INSNIYQNDDIIYT 977

Db     1846 NFALMMKSKLEKTTIRDICMYIKEDYTSRGVFLSVKFNEELIQKLFLNINAYNIKDIILK 1905

Qy      978 INSLNDYMSNTL-LHFKEKYTYPTLSTNEDIYNKEMEG------------KHIR----- 1018

Db     1906 QSHINKIKINKIHINVINKYKLHISLKNDEFIFFQMESLKKGLLDLLIYGDKDIKRCIIK 1965

Qy     1019 -------------LDDQDKY---------------DDNDNNNVDNNNKNNVDNNV----- 1045

Db     1966 KEDIEVTDNEDEICDDMDEYYNVSQGTELYERKCNSKEENKNAIRVKKEEIDDNLEKEEN 2025

Qy     1046 -----------------------DNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNF 1082

Db     2026 IIYVSEKDSVNQLKSEKKKDINDDNNNNDDNNNNNDDDNKINDTIFNDDIDSDRNNLKEN 2085

Qy     1083 NNN---------EYLSYFQK----------------NVDTIINNCLNSLDISSMYDDTKE 1117

Db     2086 GSKLENVGEHIIERLSYKMKEKNVKKEHIKKEPNLINLDTINLDTLN-------FDE--- 2135

Qy     1118 ILNNILLSKYKAEKDNVIKKYIN-EDIKNMSLEEIDK-------TAQSIYEKRKVLLTKL 1169

Db     2136 ---------------------INVHNINNEKIEFYDEHLNICQGNKKHIQKKKKKKTVYS 2174

Qy     1170 LLLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFY-----------------AY 1211

Db     2175 ILVEGNSLNYVLGLEGVDFK------HIISNHVINVFQVLGIEAARITIINEIKKCVEAY 2228

Qy     1212 SRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNKKL 1271

Db     2229 S-IDIDIRHIMLLADIMAFTGDILGINRFGI----QKARQSTLMLASFEE-----TNEHL 2278

Qy     1272 -VQDNIKNSMDHNNI 1285

Db     2279 FVSSFFKNVDEINNI 2293
```

RESULT 31
B71619
hypothetical protein PFB0280w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71619
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71619
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2539 <GAR>
A;Cross-references: GB:AE001384; GB:AE001362; NID:g3845139; PIDN:AAC71845.1; PID:g38451
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0280w

```
  Query Match             8.1%;  Score 552;  DB 2;  Length 2539;
  Best Local Similarity   20.9%;  Pred. No. 1.2e-15;
  Matches  322;  Conservative  211;  Mismatches  479;  Indels  528;  Gaps   68;

Qy      133 IYDKDNTKGI-FISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRVLTKEELI 191

Db      820 IHIKVNHKGIYYVQKKEMY--QLYDDNNNNNN----NNNKSD----------------- 855

Qy      192 LKMIDGVLLRKIGVNSQNDE----SSRSHAILNIDLKDINKNTSLGKIAFIDLAG---SE 244

Db      856 ------ICLNRVNPNKCSSEKKTNNPNSSSILKKD-KEKKKNQMDGKIVTNLVKGDNKEE 908

Qy      245 RGADTVSQNKQTQTDGAN------------INRSLLALKECIRAMDSDKNHIPFRDSEL 291

Db      909 EGNNNIIKNDDSASKGTNEHMMQRINDAETTQNNTLHKENKLCTTKDQNKIHTKINSKEN 968

Qy      292 TKVLR--------DI---------FVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNF 334

Db      969 EKVKKYYYYHINNDLGLYFYFLVGFIIKKKNCTIS--------LKLNINNLNVKYKGNNI 1020

Qy      335 KNKSTCINEEDDTNTERISILDSKG--------SEMNASSIENVVIKSNHLLSNNNNNKI 386

Db     1021 YKIKTVMYQKDIYNYYLLNILLLVGVKIYIRQHNKLNKESEYNV----------NSQNLI 1070

Qy      387 NRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKK--NKKKGLINYKSTL 444

Db     1071 GSKSKSSKIYMVHFITSE----------ISFNKKKILRPFYKIQKKINNK-----YKRII 1115

Qy      445 YNDNT-INKKHNNNNNNNNDNNDNNN------DNNNNNN----------NDSSSMVNNMI 487

Db     1116 MNQSAHINIKESKNNIISNNVEEKNSVTSNIVSNISSNNISPYYKSIKENNKMKKTNNCI 1175

Qy      488 NHMINN-----NINNNINVNNNNNNNN--------------------------------N 510

Db     1176 EHILNNYKIKYNIYEKIYIKYETNNNHMLSFKIVIDAESFSDDFFSICILFSHFILSNIN 1235

Qy      511 NN------NSHNNHLPQP----NYAFTDTSDFSSLDDMNCHLNN----------NDKSIF 550
```

```
Db     1236 ENIIFKIKNIHNQNIKESTRIYHVVFILKLFFHNLLFISCTNNSIYITKMLHPLQNIQFY 1295

Qy      551 LHKKNLR-DNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSN 609
             :||:| :| |: | :    |     | :| :: ::: :     ::: |    :|      | |
Db     1296 RYKKNIRTNNQKIYNTNYIHNKYEKIQNFVNNSKYVINDMQSLYLYVDTQNDHRIIFMST 1355

Qy      610 INKME---------DNTPKDI-----------------------LYESRNVSNMN---- 632
            |  :          || |                          |  ::|:|:|
Db     1356 ILSLIFKNIIIPKCDNVHKSFPLFFHYAKKYLHIYVQNGSNQFINTYNFQDVNNINLLHC 1415

Qy      633 ---------------------GNVLLG---------------LNKN----THHDISTK 650
                                 || ::                 :|||       : |  |
Db     1416 TKKKRPQRGSTPDEKYKGGEIKGNDIIKESDIIKCNDIIKESDVVNKNEIVENMNIIIEK 1475

Qy      651 DE------------NHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNS--IYK 695
            ||            |:| |  :    ::::: : | :::  || :    ||  |:  | |
Db     1476 DEIKTDKYTEPIKYDNTSDAKSISTSTSVLSSESSNELSDCCMNKLTKENMEMNNVIITK 1535

Qy      696 SNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNN 755
             :| | |           || : :|:|||                | || ||:|||||||
Db     1536 NNNNDNN----------NENNENNENND--------------NNENNENNDNNNNNNN 1570

Qy      756 NNIDV-------------------------------------------------ENYN 764
            ||::|                                                  :| |
Db     1571 NNVEVYKPNYKINGLQNIINSCLNFICSKRKNIKNKIKNKIIKHKKNKIINHKKKKKNCN 1630

Qy      765 NR----DGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNV----NYSMNFC 816
             |       || : |      : |  |:||:: :       |       :    :   |
Db     1631 TRHRGNTQINNKLVLINITPYILRYPNNNKSSKKLSCTKEIKKKTFPRICESYDIKKNID 1690

Qy      817 HYNLNDKNY-LID--LN-NKEQKDKNIHGCDNNIIQNRN-----DFEKKKKTNFY----- 862
             :|:| |||  ||  || :||: | :      |  |  :           | ::|: :  |
Db     1691 IHNVNKKNYKKIDDTLNVHKEEIDTSKQHTDEKICKKIQKYLYLDVKRKRYISLYMYNKK 1750

Qy      863 ---NNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSN 919
                :  ||  |             :: |  :    | ||    || || |  : |    ::| |   :
Db     1751 KGKDTNNKNIQKKKKEEEKKQISYNISSKHNSI---LNNRMKYNNIIDMYKRNNFIYKD 1807

Qy      920 NNNSSSSSSSNNNIYNNI-NDDDTFQND-----------------YCHNDN---TFTIRR 958
            :|           |  |  |:  :||                   |    |    ||  ::
Db     1808 DN-----------YKRIYTYDEILENDINISYLIKQINILNVTIICGMRNVGKTFLSKK 1855

Qy      959 -KNNTNINSNIY-QNDDIIYTINSLND-------------YMSNTLLHFKEKYTYPTLST 1003
             :||  |: :  |  |:|  :   |::|            |::  :|   :  |    |
Db     1856 IENNIIIDIDEYILKDEIKFDKLSISDFRYYEYVTFISSLYLAFYILTFDRNLSAPKDQT 1915

Qy     1004 NEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDN-NVDNNNVDNNVDNNDKN-- 1060
              | :: ::  :  :|:|   |:  :   ||||  ||  ||  |: :|||  ||    |:
Db     1916 GATIKHVDIRDEKINSKNQNKQTEYDNDINDNNNYNNSDNHNLLHNNKDNQHTSTKKKIQ 1975

Qy     1061 -----------NVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSL--- 1106
                       ||  | :|            ||:::|  : :|   |  :  |   :|  |
Db     1976 KKVSFSDVCEIYVDGPNFEN----------KNYDDNIFYTYTNKGI-TFYNKKINDLFCK 2024

Qy     1107 -----------------DISSM-------YDDTKEIL----NNILLSKYKAEKDNVIKKY 1138
                              ::: :        :| :||:|    || ||:    |: | :
Db     2025 LRKKCIQEKQNGEHQMTNVTIVLGGGIIEFDKSKEVLKKLKNTILI---KRDIDEIYDIC 2081

Qy     1139 INEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNV--DTQINN--ETSDLRKDLVM 1194
            ||::||     :::    :  :|:  :| ||       :     ||       |:   | :
Db     2082 INDNIK----PKLNGNIKDIIHRRTILYDKLSNAFHFIIPSENMINKYIRHSEYNKYI-- 2135

Qy     1195 CHICNNNP---DDQFHFYAYSRLEKDII-NLIMLRQIWCESENLRLLYQFLVVEYQNKSA 1250
                ||         |:  |   :|  :|: :|  :    :|                :  :|
Db     2136 ----NRNELIVHSFLRFFNYPFFKKPLIGDIITNYKIDKNEKN-----------DEKNDEK 2181

Qy     1251 NSVLLNVSSN--NGDIILLNKKLVQDNIKNSMDHNNIHKK 1288
            |     |  ::  |||     |      ||  |:  | ||  ||
Db     2182 NDEKNNEKNDEKNGD----NNDDNNDN--NNEDENNKKKK 2215
```

RESULT 32
T28160

```
hypothetical protein - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C;Accession: T28160; S23688
R;Kun, J.F.; Hibbs, A.R.; Saul, A.; McColl, D.J.; Coppel, R.L.; Anders, R.F.
Mol. Biochem. Parasitol. 85, 41-51, 1997
A;Title: A putative Plasmodium falciparum exported serine/threonine protein kinase.
A;Reference number: Z20482; MUID:97262159; PMID:9108547
A;Accession: T28160
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2510 &lt;KUN&gt;
A;Cross-references: EMBL:U40232; NID:g1658332; PID:g1658333; PIDN:AAB54058.1
A;Experimental source: strain FCQ27/PNG
R;Kun, J.; Hesselbach, J.; Schreiber, M.; Scherf, A.; Gysin, J.; Mattei, D.; Pereira da
Res. Immunol. 142, 199-210, 1991
A;Title: Cloning and expression of genomic DNA sequences coding for putative erythrocyt
A;Reference number: S23684; MUID:91376328; PMID:1896607
A;Accession: S23688
A;Molecule type: DNA
A;Residues: 241,'S',243-244,673-959,'R',961-977,'S',1493-1494 &lt;KU2&gt;
A;Cross-references: EMBL:X53019
C;Genetics:
A;Gene: FEST
A;Introns: 645/2
C;Keywords: surface antigen

  Query Match             8.0%;  Score 549;  DB 2;  Length 2510;
  Best Local Similarity   20.3%;  Pred. No. 1.6e-15;
  Matches  355;  Conservative  242;  Mismatches  516;  Indels  638;  Gaps   81;

Qy        2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTL-----YIDEPRYK-----------V 44
             :| :::   |     :|:   :|:    :  :|||      ::  |             :
Db       32 SSNVEIENRDNKRNEIINYRKEKHPCSNLSNMTLNPSSYHVSNPMNSHNVNVYSSSNNIM 91

Qy       45 DMTKYIERHEFIVDKVFDDTVDNFTVYEN-------TIKPLIIDLYENGCVCSCFAYGQT 97
             |  |||:             ||: :: |    |          |:   : :||    :|:
Db       92 DFRKYIK---------DDSSNDITKQSNLSSIESHTMN--LSNLYNFNNICN------- 132

Qy       98 GSGKTYTMLGSQPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKL-YD 156
                         ||   |                   ||||    :    || : :|
Db      133 ---------EINSYGFKSNP-----------------LDNTK--VQNMNSSYCNQMKYS 163

Qy      157 LLQKRKMVAALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLR-----KIGVNSQNDE 211
             :             ||          |           |      |    ||  :|   |
Db      164 DYYQHKPCLCHRRG----------ILNTSNSSENQINSSDVCSLHKNDYTKISMNFPNAS 213

Qy      212 SSRSHAILNIDLKDINKNTSLGKI---AFIDLAGSERGADTVSQNKQTQTDGANINRSLL 268
             ||     :|:|     | | :    :    : ::     :|:| ::  :  |   :  ||   |
Db      214 SS-----VNVDSNVYNSNITAPCVMNSSPMNYNNGKWGSDIINMMQTTNMNNLKINDLYL 268

Qy      269 ALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYS 328
                      .::: | |  |:||:|                          ||      :
Db      269 ----------NEINKINFQDSDL-----------------------SICSHGTS----- 289

Qy      329 SRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNN--NNNKI 386
                |: |:||          | :|     :||  :                :| ||  ||:  :|| :
Db      290 ---KSKKHKS-------DYSTVPYNILSNN-------------MKPNHFTSNSVPSNNIL 326

Qy      387 NRGKINDKIERNN-----ILKNK--SFDKPREGFTST--FGKYSSLNDIDKIKKNKKKGL 437
             :       |  |   ||   |:  |    :    |  :      ||    :  :|   |: :  :
Db      327 SNNMIPNNIALNNMSSSSIISNNMASCNIPLNNMISTMEYQQYILSNLLLKVYQQHTSTY 386

Qy      438 INYKSTLYNDNTINKKHN------NNNNNNNDNNDNNNDNNNNNNNNDSSSMV-------- 483
             ::|   :|      ||   :|       || :||   :| ::|    | :|| ||: :
Db      387 VSYDMLIYEHMLINMSNNMSSNMSSNMSSNMSSNMSSNMARNVSNNMSSNFISSNLSNSF 446

Qy      484 ----------------NNMI-NHMINNN------INNNINVNNNNNNNNNNNSHNNHLP 520
                                 || | :|:| ||       |:||:|:||  :||   :|:     |:
Db      447 ISNNMSSNLFLKSLISNNFIPDHLIPNNSNSKNIISNNVNLNNFVSNNFTSNHMEYNNRK 506

Qy      521 QPNYAFTDTSDFSSLDDMN------------CHLNNNDKSIF--LH--KKNLRDNIKLK 563
             : |    :    :   : ::||                |   :|:  :|   ||    :    :|| ::|
```

```
Db       507 KKNLIASSKVNTLNNENMNSLSEKLNPLILSSSHSIHNENQVFNNLHFSEHMQKDNEEIK 566
Qy       564 NRSSCDNIMNK--KK-------NNLHL-----ARHSVGSKLTMFSYDP----------- 597
Db       567 NDKVIKHMQEEQIKKTPEAYLLNNYYLMFQQKLENQQSYTKNMKTDDPIKFNDIIINQQN 626
Qy       598 -------------QKNK----DNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVL--- 636
Db       627 ICNSSMKNYNNINEEKNKMCSHENITENMNISRIKNNDDKQIV-----VSKRSSNVLRQM 681
Qy       637 ---------------------------------------------------------LG- 638
Db       682 IDFPSVNDKIQNESFANTRRKAREMNIQNDICSEAYKSETIKDRLRSRARKKGLEGVVGG 741
Qy       639 ---------------LNKNT-------------------------HHDISTKDENHN 655
Db       742 RTTEANSQSLSKYDFLLNKNTIPNNTINLRNKKESKQNAQINNSKETCYHIIEDIDSNVN 801
Qy       656 D--------NKI------------------------NNGVINIINNSNVNSINNSN--- 679
Db       802 EQITKDIDMNEIKSSLKVDENMMRFLCADEKKEEMNQNHEQENILDCANKKVFNNKNGSS 861
Qy       680 ----MNSINN-----------SNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKN--N 722
Db       862 RRKRNNEIKNKKHEEGKDKEDKNLNDKDMEKTNVEEN----DIEEENITEENITEENVTE 917
Qy       723 DNIFFDAISCDNNMYPNITNNN-NNNNNNNNNNNNNIDVEN-----YNNRDGTNNSMKLY 776
Db       918 ENITEENVTEENITEENITEENVTEENEDKNSDDKNVEERNCLASAHGRRFIQTRSRYKQ 977
Qy       777 AYN-SHNLFQPDNNK---NTSNIQNINTNKNNQDGNVNYSMNFCHYNL------NDKNYL 826
Db       978 LYNEKHRINKNMHNKQINKMSQINNRNRNMNNMDEEKKSSTDICDHSREMENDENEENVE 1037
Qy       827 ID----LN---NKEQKDKNIHGCDN------------NIIQNRNDFEKKKKTNFYNNNNI 867
Db      1038 LDNIKELNGIINKE----NIENDENIKPDMKKNKEKVNVITNERNKRKREKTKCTDE--- 1090
Qy       868 VIVNNNMGNNNS----PRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNS 923
Db      1091 ---KNDQGKDNTHEHMVEKKVKLEEDSEEMERNTNNNKINYDKNNNDEKNNNTDCTNCTY 1147
Qy       924 SSSSSSNNNIY---NNINDDDTFQNDY------CHNDNTFTIRRKNNTNINSNIYQNDDI 974
Db      1148 CEDNIKENNIIFLNGNVENIDNDKNKLVSETVECTND----INEKKKIGEQIACTNNDKM 1203
Qy       975 IYTINSLND---------YMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKY 1025
Db      1204 LYSVNVINDDFNMKNVVEFLSHNDVAIKKELNY------LNIENKIHEREYI-LDNKKTL 1256
Qy      1026 DDNDNNNVD---NNNKNNVDNNVDNNNVDNNVDNNDKNNVDN------------------ 1064
Db      1257 HIDNKKNKDIYYNIAKSGIDENMINSNLFDNYDN--KNNVRDLPRDLYSSTDGTMDIIIN 1314
Qy      1065 ------------NNVDNDDDDVDFHNIKN--FN---NNEYLSYFQKNVDTI--------- 1098
Db      1315 KKKKIWLNIGCINYMDNFVGKGSYGSVRKVLYNIDLNNEYLKFYMDKINDVKFLKVLTHT 1374
Qy      1099 ------INNCLNSLDISS----------MYDDTKEILNNILLSKYKAEKDNVIKKYINED 1142
Db      1375 CHMKKGLKYLVDKYDIASDSDCVSYWNDYYLELKDHLDDEKYEEITKETKNVQKV---QD 1431
Qy      1143 IKNMSLEE-IDKTAQSIYEKRKVLLTKLLLL---------------FK---KNVD-TQIN 1182
Db      1432 EKNKDNEKSINKNTNKLNEKELVILESDLMLNNKLDNCLCCHKKRYFELAVKEIDISRKG 1491
Qy      1183 NETSDLRKDLVMCHI-CNNNPDDQFHFYAYSRLE------KDIINLIMLRQIWCESENLR 1235
Db      1492 NEFKFLREQELLCHFNCN----------VIKPLSTKIGHLKDKQNYEIL--MHCATGDLR 1539
Qy      1236 LLYQFLVV----EYQNKSANSVLLNV-------------SSNNGDII-LLNKKLVQDNIK 1277
Db      1540 KLLQNLLYHRKKEYEKKKVNKILKLIHICFGRKYHCFESLFNLECIGLCYKKL--KNMK 1597
Qy      1278 NSMDHNNIHKK 1288
Db      1598 MKINNNTIEIK 1608
```

RESULT 33
T18472
hypothetical protein C0440c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18472
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z18937
A;Accession: T18472
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2269 <LAW>
A;Cross-references: EMBL:AL008970; NID:e1407852; PID:e1332566; PIDN:CAA15615.1
C;Genetics:
A;Gene: C0440c
A;Map position: 3

```
  Query Match             8.0%;  Score 544;  DB 2;  Length 2269;
  Best Local Similarity   21.4%;  Pred. No. 2.3e-15;
  Matches 273;  Conservative 169;  Mismatches 327;  Indels 504;  Gaps   65;

Qy       341 INEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNI 400
Db         1 MNNTFKINKKRRTVYES-------YNIINIANKKRYMC---NDNKMD----ND----NNI 42
Qy       401 LKNKSFDKPREGFTSTFGKYSSLNDIDKIKKN----KKKGLINYKSTLYNDNTINKKHNN 456
Db        43 INEKKYFNNSSSIKNTGNNYKN----DNIDKSYDMCERRSMMDRDENIYNNNYKNKNRNN 98
Qy       457 -NNNNNDNDNDNNNDNNNNN-NNDSSSMVNNMIN-HMINNNINNN------------IN 500
Db        99 YSNKNNYSNNINHMDNRRNTYKGDLDGRKQYLSNKHYESKNDNENSYNLKKYNFNIPPIK 158
Qy       501 VNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNI 560
Db       159 YNNNNNNNNNNNNDRQNIDNNIYKFYD----------------DK------------- 188
Qy       561 KLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKD 620
Db       189 ------------NEKKSN---KTHSYGNNF--------NNEFNTFNDPKQNYM-SNESHN 224
Qy       621 ILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSN- 679
Db       225 MTQDGHPIYNHNNNNNDGRIKN-RKDISL--SSYGYNKISNDNLSKENTSYINQYNNSGG 281
Qy       680 -----MNSINNSNMNSNSIYKSNYN--SNQSISDVQIRYVNEMDTSNKNNDNIFFDAISC 732
Db       282 ANRHVINAQHNNNQHNNNQHNNNFNYHNYKNAKPYQDKY--------KNGDSFKFSAYNA 333
Qy       733 ------DNNMYPNITNNN------------------------------------------ 744
Db       334 EQHDNMDKNYYQDIHKNNMKHSTKGDTAISFNGQKVFGSFKNKQMKEKRKFFKNKERGSS 393
Qy       745 -------NNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPD----NNKNTS 793
Db       394 INRPGMHNNNNNYNNNNNNY----YNN------------YNNH--FCDDAERRNNKN-E 434
Qy       794 NIQNINTNK-NNQDGNVNYSMNFCHYNLNDK---NYLIDLNNKEQKDK--------NIHG 841
Db       435 NWCEINSNELYEKDGDMPI------YNANENCNLKKMIYKNDKKEKEKIKILLTEDNVNN 488
Qy       842 CDNN-----IIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNS----------------- 879
Db       489 VQSNIEKLSIYKSRNEI-----IEMIEKNDVTFINGETGSGKSTCVPKFLLEENIRENKK 543
Qy       880 -------PR---------------------MKYGLCG--------------------- 888
```

```
Db     544 INIIVTEPRRIACIALSKILSELTNEKLGQKIGYRISGESLYDSEKTVITYITIGYLFKL 603

Qy     889 -----------SHTSIDNMKNNEM------------KNNEMKDNEM-------------- 911

Db     604 FLHHKNMYKKFTHVIIDEIHDRSILLDIVLLFIKLYLHNKQKDEQMFKLIIMSATMQSNL 663

Qy     912 ----------------------------------------KDNHIKSNNNNSSSSSSNN 931

Db     664 FYSYFEHPNIKMGSIFIGTKIFSIDTFYIEDIINYTRYGSRNIHDEKINDEKINDEKIND 723

Qy     932 NIYNN--INDDDTFQND---------YCHNDNTFT---------IRRKNNTNINSNIYQND 972

Db     724 EKINDEKINDDNKKSCDDKKSCDDKKSCDGNNNFVKESVIDFILRRKNCNKI--NLSKNS 781

Qy     973 D-IIYTINSLNDYMSNTLLHFKEKY-----------------TYPTLST--NEDIYNKEM 1012

Db     782 EMLLYKIKSEYDKNIHMFNNNNNKYCNDKNKDLDVDEIIPANVFSNISNLCLELVYNLCL 841

Qy    1013 EGKHIRL------DDQDKYDDND---NNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVD 1063

Db     842 KGDSVIIFLSGMQDITDMYHQLSMIINNNMDNPSNSNI-------NINNGNDANDANDVN 894

Qy    1064 N-NNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDT---IINNCLNSLDISSMYDDTKEIL 1119

Db     895 NMHNMHN------MHNMHNIANQ--LTYHRTDIKIHIHMLHSCL--------YDNTIHKL 938

Qy    1120 N------NILLSKYKAEKDNVIK--KYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLL 1171

Db     939 KHNDTDINIFLSSNIAESSITIPNVRLVIDFCIQKNIEYNDK------KKAHILVKK--- 989

Qy    1172 LFKKNVDTQINNETSDLRKDLV--MCH-ICNNNPDDQFHFYAYSRLEKDIINLI---MLR 1225

Db     990 --------WINKSSMEQRKGRCGRTCHGIC------------IRMISKNFLNLLRDHKIS 1029

Qy    1226 QIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDII------LLNKKLVQDN---- 1275

Db    1030 EIYTHS--LHLLYLYIL------KSMSVLNGLINKRNEVVHCDNTNVYNVKEVKDQKNGI 1081

Qy    1276 IKNSMDHNNIHKK 1288

Db    1082 CKNEISHSHAENK 1094
```

RESULT 34
A44396
P-type cation translocating ATPase - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jun-2000
C;Accession: A44396
R;Krishna, S.; Cowan, G.; Meade, J.C.; Wells, R.A.; Stringer, J.R.; Robson, K.J.
J. Cell Biol. 120, 385-398, 1993
A;Title: A family of cation ATPase-like molecules from Plasmodium falciparum.
A;Reference number: A44396; MUID:93132070; PMID:8421054
A;Accession: A44396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1984 <KRI>
A;Note: sequence extracted from NCBI backbone (NCBIP:122678)

```
  Query Match             7.9%;  Score 543;  DB 2;  Length 1984;
  Best Local Similarity   21.5%;  Pred. No. 2.2e-15;
  Matches  321;  Conservative  202;  Mismatches  412;  Indels  560;  Gaps   73;

Qy      22 KDSDIITVKNN--CTLYIDE-------PRYKVDMTKYIERHEFIVDKVFDDTVDNFTVYE 72

Db     568 KDDNMIHNKNSIISEYYIKDNMKNLHTSSKKKSITK--ERSNFLV-----QTIKSCLLKD 620

Qy      73 NTIKPLIIDLYENGCVC--------SCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAA 124

Db     621 HYIKEKKKEYYTNNTYCNDLHINDSTCSSY----------LLNSE--------------- 655

Qy     125 GDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRV 184

Db     656 --------------TKDAYCEYYNI--DHLCDINKKNMDI----NSKNELMGK------- 688

Qy     185 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAG-- 242

Db     689 YSKNELMGKTIKNELMGKYSKNELMGKYSKNEL-----MGKYSKNELMGKYSKNELMGKY 743

Qy     243 --SERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300

Db     744 SKNELMGKTIKNQVGVDTNIYHMN------------CDNDYNY----------------- 774

Qy     301 GKSKSIMIANISPTISC--CEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358

Db     775 ---------DYPCDYNCNNCNDTYHRLEYHN-----------INKDNSFN-----IPPEK 809

Qy     359 GSEMNASSIENVVIKSN-----------HLLSNNNNNKINRGKIN---------DKIERN 398

Db     810 NKSYNNISEH---IKINYPLLFEALACCHTLS-KVNNKIMGDVLEILMFNFTNCDMLINN 865

Qy     399 N--ILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKG-----------LINY----- 440

Db     866 NSFIIKEK--------------KKNCSYDFQKIDGDKNIGANDERCHLNNNLVSYNILKR 911

Qy     441 ---KSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSMVNNMINHMINNNINN 497

Db     912 FEFQSRLQRMSVIVKSTYGNNNDDNNDDDNNNDDDNNDDN-----------------NN 953

Qy     498 NINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLR 557

Db     954 DDNNNDDNNDDNNNNNYYYN------------------------------IFC--KGSP 980

Qy     558 DNIK---LKNR--SSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKD-------NTF 605

Db     981 EKIKELCLKSKIPNNYDEILNKYTKQ--------GMRILSISYKRVKSKNINLLNVKRSF 1032

Qy     606 FKSNI---------NKMEDNTPKDILYESR-----------------------NVSNMN 632

Db    1033 VESNLHFLGFLIFTNNMKKNAP-DIIHNLQTSGCQCIMSTGDNVLTSIHVAKKCGIINSN 1091

Qy     633 ------GNVLLGLNKNT---------HHDISTKDENHNDNKINNGVINIINNSNVNSINN 677

Db    1092 VESIIIGDVIPVVGKNNKQKKKLVWYNHKNDTYLKGHDKTCIDNEFTSIQSQMNSDNICG 1151

Qy     678 SNM--NSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDN- 734

Db    1152 DNICGDNIYGDNICGDNIYGDNINGD-----------------NINGDNIYGDNINGDNI 1194

Qy     735 -----NMYPNITNNNNN-----------NNNNNNNNNNIDVENYNNRD-GTNNSMKLYA 777

Db    1195 NGDNINTYDNIYGDNYNLDYCPTEYHKCTYNNSILYRNNFLYKKENKKDKNYKNISTLYE 1254

Qy     778 YNSHNLFQPDN------NKNTSNIQNINTNK---------------------------- 802

Db    1255 HRTNDI-QFDKLCDILINKDPRNVNIVLTGKAFIFLKKKFYSFHLPYYEECKNIVHYIMK 1313

Qy     803 ----------NNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQK--DKNIHGCDNNIIQN- 849

Db    1314 KKHKKIKNIINNHNSNLYY-----HYNIIDT--FVKRMNKEYMCFNKLLYKIQQKLLYNL 1366

Qy     850 -RNDFEKKK-KTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMK 907

Db    1367 IHNLYKKKKYMNNYYDIDEVHLIGNNNNNNN-----------------KNNSKEKKLPL 1408

Qy     908 DNEMKDNHIKSN--NNNSSSSSSNNNI----YNNINDDDTFQNDYCHNDNTFTIRRKNN 961

Db    1409 KNKMK--HIRKNESNDNITFNTYTSNNIHLSKYKYVHHKNYYYPDSCTN-----LRKKKN 1461

Qy     962 T---NINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDI-YNKEMEGKHI 1017

Db    1462 SLFYNLKKYIYYEK---------KKYLQHCLLK-HDNYKKVELPRIKDINYSYQMESIKT 1511

Qy    1018 R-----LDDQDKYDD---------NDNNNVDNN----NKNNVDNNVDNNNVDNNVDNNDK 1059
```

```
Db       1512 RNFIHSLSEQFAFSNLILSFYIIKNDDNNVYNKNYIYNKNYIYN---KNSICNKNYICNK 1568

Qy       1060 NNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEIL 1119
              | : | |             || | | ||      | :||: |   :  |     :|        :
Db       1569 NYIYNKN-----------NIYNKNN----IYNKKNILTHAKSVL----LSGSSKKFLKFF 1609

Qy       1120 NNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDT 1179
              :||:        || |       ::||   :   :: |:      |  ::|     |||:  :
Db       1610 SNIIRRHKLKEKKN------KKNIKRYKMNHVNNTS-----KGHIILNMCTHGFKKDYSS 1658

Qy       1180 Q------INNETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESEN 1233
                     :||:    |: | |                   | |    :: ::    |   |
Db       1659 LKNKYRIVNNKRYMLKNDNV-----------------YDRHMYNLTDMYRGTQYGCSKK- 1700

Qy       1234 LRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNK------KLVQDNIKNSMDH 1282
                             :||        |:  ||  :  || ||        |:  |   | :: |
Db       1701 ------------KNK-------NIYMNNNNNILKNKINRFLEHLLVDKCKRNICH 1736
```

RESULT 35
T18502
hypothetical protein C0755c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18502
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18502
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1553 <LAW>
A;Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331923; PIDN:CAB11141.1
C;Genetics:
A;Map position: 3
A;Note: C0755c

```
  Query Match             7.9%;  Score 542.5;  DB 2;  Length 1553;
  Best Local Similarity   24.4%;  Pred. No. 1.8e-15;
  Matches  243;  Conservative  162;  Mismatches  307;  Indels  283;  Gaps   54;

Qy        341 INEEDDTNTERISILDSKGSEMNASSIENVVI---KSNHLLSNNNNNKINRGKI----ND 393
              :| : : | |: |  : :    ||     :|:::    |:: ::    |   |:   ||     |
Db          1 MNIDQNNNIEK-KIANKRKGNMNKK--KNILLNQPKNDEVIMKKNMKKVKNEKICKNGKD 57

Qy        394 KIE----------RNNILKN---KSFDKPREGFTSTFG---KYSSLNDIDKIKKNKKKG 436
               ||          ::| :|     |  :| ::  :|| |   | | | |::| ::: :||
Db         58 NIEETSTHLINRRRKDNHIKEAIYKDLEKEKKFASSTKGTSIKSSGLLDLNK-EEHVEKG 116

Qy        437 LINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNI- 495
              ::: || :     :: |        |       |: | |   ||||||:      :||: |    |||
Db        117 MVDNKSVITRTSS-NYSILNYFKNSKDTNKSGMTNNNNNNNN----INNINN---NNNIV 168

Qy        496 ------NNNINVNNNNNNNNNNNNNSHN---------------NHLPQPNYAFTD----- 528
                    |  |::|| || :|  | :|                   |: : |    |
Db        169 KTSSGSNKTRNISNNRNNIHNKPNGYNLKRDNIKITNYMKQSDRHIEKNNEVHLDKHGYK 228

Qy        529 TSDFSSLDDMNCHL---NNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHS 585
                ::     : | :|    || : :: :||    | |   :| || |: :|   |: :
Db        229 DDNYKKTFNHNNYLSMKNNIENNLMNYKKCKLDRIHAENNSSIDSFQSKDDKNVIIE--- 285

Qy        586 VGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVS-NMNGNVLLGLNKNTH 644
                      : |   |||:     |  :|               | | :| | :|  |  || : |
Db        286 --------NKDTYKNKEYMINKDVVN----------LTEHKNDSYDMVCN-RLGTSCNVL 326

Qy        645 HDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSN--YNSNQ 702
              : ||    :| ||| ||   :|  |||  |                :  | :  |  |  ||  :
Db        327 NTISLSQKN-NDN-IN---LNACNNSLVIKGEERKSRCTGQNRASSVGLLKRNSIYNYKE 381

Qy        703 SISDVQIRYVNEMD----TSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNN------- 751
              :: |   |    ||:     | | ||         : | :|:   |  || |||| |
Db        382 NLRDDLINNCVEMENMDTTKNINN----MKNLDCVSNI--NYVNNINNNVNINKGLINSS 435

Qy        752 ---NNNNNIDVE-NYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNIN----TNKN 803
                 ||:  ||: | |  |::  ||:    : ||             :||  : :||     :| |
Db        436 QEINNSCKNIEYELNDLNKEEENNN---FLYNF-------KERNTEYLHSINIPYTSNSN 485

Qy        804 NQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYN 863
              |    |:|              ||:|:      | ::|:     |        :|  |||       |
Db        486 N---NIN-----------KNHLMTSLPTEYRNKSSKSSDELFSRNLLDFENFCSYKFKR 530

Qy        864 NNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNS 923
              |    ,  :|||: ||            :|   :   ::::     ::       :||    |    ::    :||
Db        531 N-----INNNLVNN--------ICNMYEEVNDL---DVYPEQMKRGEGSVSY--GDNNMC 572

Qy        924 SSSSSSNNNIYNNINDDDTF----QNDYCHNDNTFTIRRKNNTNI--NSNIYQNDDIIYT 977
              :: :    |||||  |   :   |      ::|:     :    |     |       |:    |  | |  |::: :
Db        573 NNRNGYENNIYNTIKRNSYFFHPYKDDHFEGEKLFKKPRICVYNVINNGNKYDNNNLSVS 632

Qy        978 ----------IN-------SLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEME------- 1013
                        :|                |: :  |: |:: :||          :   |||: |||:
Db        633 HYDDVEKRRRVNLGSSGNMDLHYHHSDLLINKREK-----VIINEDVNNKEIMKGYIKNF 687

Qy       1014 ---------GKH-----------------IRLDD-------QDKYDDNDNNNVDNNNKN 1039
                       ||:                   |  |:              :|  :    :  |  :
Db        688 WYLSKKIFFGKYKNCAVSLENEEVERLKEIISFDEKKGKYTLEDLFGWEERKNFERKKEE 747

Qy       1040 NVDNNVDNNNVDNNVDNNDKNN------VDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQK 1093
               | :     |  :  |   |  | ::|              :|||  :||||            :::|:
Db        748 KKDTH-GGNKMGNYGDKNWEDNYCKSEYYNNNNNNNDDDDA-------YDDND------- 792

Qy       1094 NVDTIINNCLNSLDISSMYDD-TKEILNNILLSKYKAEKDNVIKKYINEDIKNMSLEEID 1152
                               | |:: |:   |:|  :: :|      ||| |   |        :||   |: ||
Db        793 ------------DDSTLLDEGMKDICDDETIS----EKDYVTDK----KLKNFRLDLID 831

Qy       1153 ---KTAQSIYEKRKVLLTKLLLLFK---KNVDTQI 1181
                    ||:||:   :    |:   :           ||  |   |
Db        832 GFLYDKQSLYEREMIENEKIFSMISFNHKNYDIHI 866
```

RESULT 36
S14032
kinesin-related protein cut7 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Feb-2001
C;Accession: S14032
R;Hagan, I.; Yanagida, M.
Nature 347, 563-566, 1990
A;Title: Novel potential mitotic motor protein encoded by the fission yeast cut7(+) gen
A;Reference number: S14032; MUID:91015362; PMID:2145514
A;Accession: S14032
A;Molecule type: DNA
A;Residues: 1-1073 <HAG>
A;Cross-references: EMBL:X57513
A;Note: the authors translated the codon GCC for residue 2 as Gly and AGT for residue 1
C;Genetics:
A;Gene: cut7
A;Introns: 50/1
C;Function:
A;Description: probably involved in mitotic spindle body duplication
C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
C;Keywords: ATP; microtubule binding; mitosis; nucleotide binding; P-loop
F;61-415/Domain: kinesin motor domain homology <KMOT>
F;147-154/Region: nucleotide-binding motif A (P-loop)
F;153/Binding site: ATP (Lys) #status predicted

```
  Query Match             7.9%;  Score 541.5;  DB 2;  Length 1073;
  Best Local Similarity   20.8%;  Pred. No. 1.3e-15;
  Matches  236;  Conservative  222;  Mismatches  410;  Indels  265;  Gaps   43;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYID-------EPRYKVDMTKYIERHEF 55
               : | ||||  |  :: |            |::| :| :         |    : | :   :    :
Db         59 TNINVVVRVRGRTDQE----------VRDNSSLAVSTSGAMGAELAIQSDPSSMLVTKTY 108

Qy         56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD 115
                ||||      |     ::||:: |::   :   ||   |: |||||||:||||||  |        ||
```

```
Db        109 AFDKVFGPEADQLMLFENSVAPMLEQVL-NGYNCTIFAYGQTGTGKTYTMSGD----LSD 163

Qy        116 TPGIFQYAAG----DIFTFLNIYDKDNTK-GIFISFYEIYCGKLYDLL---QKRKMVAAL 167

Db        164 SDGILSEGAGLIPRALYQLFSSLDNSNQEYAVKCSYYELYNEEIRDLLVSEELRKPARVF 223

Qy        168 EN----------GKKEVVVKD----LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESS 213

Db        224 EDTSRRGNVVITGIEESYIKNAGDGLRLLR-----------EGSHRRQVAATKCNDLSS 271

Qy        214 RSHAILNIDLK--------------DINKNTS----LGKIAFIDLAGSERGADTVSQNKQ 255

Db        272 RSHSIFTITLHRKVSSGMTDETNSLTINNNSDDLLRASKLHMVDLAGSENIGRSGAENKR 331

Qy        256 TQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTI 315

Db        332 ARETGM-INQSLLTLGRVINALVEKAHHIPYRESKLTRLLQDSLGGKTKTSMIVTVSSTN 390

Qy        316 SCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISI------LDSKGSEMNASSIEN 369

Db        391 TNLEETISTLEYAARAKSIRNKP----QNNQLVFRKVLIKDLVLDIERLKNDLNATRKKN 446

Qy        370 VV-------------IKSNHLLSNNNNNKINRGKINDKIERNNI--LKNKSFDKPREGFT 414

Db        447 GVYLAESTYKELMDRVQNKDLLCQEQARKLEVLDLNVKSSREQLQYVSKSNQEHKKEVEA 506

Qy        415 STFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNN------------ 462

Db        507 LQLQLVNSSTELESVKSENEK----LKNELVLEIEKRKKYETNEAKITTVATDLSQYYRE 562

Qy        463 ---------DNNDNNNDNNNNNNNDSSSMVNNMINHM--INNNINNNIN-----VNNNNN 506

Db        563 SKEYIASLYEKLDRTERNNKENENNFWNLKFNLLTMLRSFHGSFTDETNGYFTLLNDFNA 622

Qy        507 NNNNNNNSHNNHLPQPNYAFTD-TSDFSSLDD------MNCHLNNNDKSIFLHKKNLRDN 559

Db        623 SMEELLNTHSNQL---LISMTKITEHFQSLDEALQSARSSCAVPNS--SLDLIVSELKD- 676

Qy        560 IKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPK 619

Db        677 ----SKNSLLDALEHSLQDISMSSQKLGNGISSELIELQKDMKESY-------------R 719

Qy        620 DILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSN 679

Db        720 QLVQELRSLYNL---------QHTHEE--------SQKELMYGVRN-----DIDALVKTC 757

Qy        680 MNSINNSNM---NSNSIYKSNYNSNQ-----SISDVQIRYVNEMDTSNKNNDNIFFDAIS 731

Db        758 TTSLNDADIILSDYISDQKSKFESKQQDLIANIGKIVSNFLQE------QNESLYTKA-- 809

Qy        732 CDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKN 791

Db        810 -------DILHSHLNDTNSNIRKANEI----------MNNRSEEFLRNA----------- 841

Qy        792 TSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCD------NN 845

Db        842 ASQAEIVGANKERIQKTVENGSQL----LDSKSKAIHSNSRSMYDHCLALAESQKQGVNL 897

Qy        846 IIQNRNDFEKKKKTNFYNNNN------IVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNN 899

Db        898 EVQTLDRLLQKVKEHSEDNTKEKHQQLLDLLESLVGNNDN------------LIDSIKTP 945

Qy        900 EMKNNEMKDNEMKDNHIKSNNNNS--SSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIR 957

Db        946 HTELQKITDHVLKGTTSLANHTNELLGLGDESLCNLETTIEDTSLVK-----LETTGDTP 1000

Qy        958 RKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNK 1010

Db       1001 SKRELPATPSWTRDSSLIKETTNLN--LDSDKKFVRETYTSSNQSNEPDVYDK 1051
```

RESULT 37
T38378

kinesin-like protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C;Accession: T38378
R;McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21790
A;Accession: T38378
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1085 <MCL>
A;Cross-references: EMBL:Z70691; PIDN:CAA94636.1; GSPDB:GN00066; SPDB:SPAC25G10.07c
A;Experimental source: clone c25G10
C;Genetics:
A;Gene: SPDB:SPAC25G10.07c
A;Map position: 1
C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

```
  Query Match             7.9%;  Score 538.5;  DB 2;  Length 1085;
  Best Local Similarity   20.7%;  Pred. No. 1.8e-15;
  Matches  235;  Conservative  223;  Mismatches  410;  Indels  265;  Gaps   43;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYID-------EPRYKVDMTKYIERHEF 55

Db         71 TNINVVVRVRGRTDQE----------VRDNSSLAVSTSGAMGAELAIQSDPSSMLVTKTY 120

Qy         56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD 115

Db        121 AFDKVFGPEADQLMLFENSVAPMLEQVL-NGYNCTIFAYGQTGTGKTYTMSGD----LSD 175

Qy        116 TPGIFQYAAG----DIFTFLNIYDKDNTK-GIFISFYEIYCGKLYDLL---QKRKMVAAL 167

Db        176 SDGILSEGAGLIPRALYQLFSSLDNSNQEYAVKCSYYELYNEEIRDLLVSEELRKPARVF 235

Qy        168 EN----------GKKEVVVKD----LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESS 213

Db        236 EDTSRRGNVVITGIEESYIKNAGDGLRLLR-----------EGSHRRQVAATKCNDLSS 283

Qy        214 RSHAILNIDLK--------------DINKNTS----LGKIAFIDLAGSERGADTVSQNKQ 255

Db        284 RSHSIFTITLHRKVSSGMTDETNSLTINNNSDDLLRASKLHMVDLAGSENIGRSGAENKR 343

Qy        256 TQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTI 315

Db        344 ARETGM-INQSLLTLGRVINALVEKAHHIPYRESKLTRLLQDSLGGKTKTSMIVTVSSTN 402

Qy        316 SCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISI------LDSKGSEMNASSIEN 369

Db        403 TNLEETISTLEYAARAKSIRNKP----QNNQLVFRKVLIKDLVLDIERLKNDLNATRKKN 458

Qy        370 VV-------------IKSNHLLSNNNNNKINRGKINDKIERNNI--LKNKSFDKPREGFT 414

Db        459 GVYLAESTYKELMDRVQNKDLLCQEQARKLEVLDLNVKSSREQLQYVSKSNQEHKKEVEA 518

Qy        415 STFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNN------------ 462

Db        519 LQLQLVNSSTELESVKSENEK----LKNELVLEIEKRKKYETNEAKITTVATDLSQYYRE 574

Qy        463 ---------DNNDNNNDNNNNNNNDSSSMVNNMINHM--INNNINNNIN-----VNNNNN 506

Db        575 SKEYIASLYEKLDRTERNNKENENNFWNLKFNLLTMLRSFHGSFTDETNGYFTLLNDFNA 634

Qy        507 NNNNNNNSHNNHLPQPNYAFTD-TSDFSSLDD------MNCHLNNNDKSIFLHKKNLRDN 559

Db        635 SMEELLNTHSNQL---LISMTKITEHFQSLDEALQSARSSCAVPNS--SLDLIVSELKD- 688

Qy        560 IKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPK 619

Db        689 ----SKNSLLDALEHSLQDISMSSQKLGNGISSELIELQKDMKESY-------------R 731

Qy        620 DILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSN 679

Db        732 QLVQELRSLYNL---------QHTHEE--------SQKELMYGVRN-----DIDALVKTC 769
```

```
Qy        680 MNSINNSNM---NSNSIYKSNYNSNQ-----SISDVQIRYVNEMDTSNKNNDNIFFDAIS 731
              |:|::::   :  |  || :  | |      :|  :  :: |       |:::: |
Db        770 TTSLNDADIILSDYISDQKSKFESKQQDLIANIGKIVSNFLQE------QNESLYTKA-- 821

Qy        732 CDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKN 791
                     :| :::  |: |:|     | |           ||  :  :  |:
Db        822 -------DILHSHLNDTNSNIRKANEI----------MNNRSEEFLRNA----------- 853

Qy        792 TSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCD------NN 845
                |  : :  ||      |           |: |:  |  |::   |   :    :      |
Db        854 ASQAEIVGANKERIQKTVENGSQL----LDSKSKAIHSNSRSMYDHCLALAESQKQGVNL 909

Qy        846 IIQNRNDFEKKKKTNFYNNNN------IVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNN 899
               :|  :    :|  |  :  :|           :  :: :  :|||::           ||::|
Db        910 EVQTLDRLLQKVKEHSEDNTKEKHQQLLDLLESLVGNNDN------------LIDSIKTP 957

Qy        900 EMKNNEMKDNEMKDNHIKSNNNNS--SSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIR 957
                :  :: |: :|       :|: |           |  |:    |   |    :    |
Db        958 HTELQKITDHVLKGTTSLANHTNELLGLGDESLCNLETTIEDTSLVK-----LETTGDTP 1012

Qy        958 RKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNK 1010
               |       :   ::  :|      :||   :  :       :| ||     :     |:|:|
Db       1013 SKRELPATPSWTRDSSLIKETTNLN--LDSDKKFVRETYTSSNQTNEPDVYDK 1063
```

RESULT 38
E71609
Ser/Thr protein kinase PFB0665w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: E71609
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: E71609
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1714 <GAR>
A;Cross-references: GB:AE001409; GB:AE001362; NID:g3845241; PIDN:AAC71922.1; PID:g384524
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0665w

```
  Query Match             7.9%;  Score 537.5;  DB 2;  Length 1714;
  Best Local Similarity   20.8%;  Pred. No. 3.3e-15;
  Matches  297;  Conservative  190;  Mismatches  373;  Indels  569;  Gaps   66;

Qy        324 TLRYSSR-----------VKNFKNKSTCINEED------DTNTERISILDSKGSEMNASS 366
               | || :            : | |||   | | :       :  | :|:|: | |   : : ::
Db         20 TKEYSKKRFDKFNDIYEIITNHKNKQPHIKENNIKYITRNVNYDRLSV-DEKKKKNDINN 78

Qy        367 IE--------NVVIKSNHLLSNNNNNKI-NRGKINDKIER-NNILKNKSF----DKPREG 412
               |:         : |: : |   |::|||: :  |   |   |  | | |  |       :: :|
Db         79 IDKYEKTKTCSYVLNNLHKKYNHHNNKMYDEYKFYDYYELINKIKKLKGFKNVIEERGKG 138

Qy        413 FTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNN 472
                 :   |  |:  ||        |||                 |||  ||||||::||| ||| |
Db        139 NDNRLGVSSTSND-------KKKN--------------NKKRYNNNNNNDNNNDINNDCN 177

Qy        473 NNNNND----------SSSMVNNM-----------INHMI-----NNNINNNINVNN-N 504
               || |              ||| ||             :::        ||  : :  |: : |
Db        178 NNKYNPCCSSCNGNVLSSSKTFNMCEGDKKISYGRQITNLVSCYKYNNQLKSPYNIHTIN 237

Qy        505 NNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKN 564
                  ::||     | |:  ||  ||    :|: : |        ||  |   ||:|
Db        238 QQVHDNNIYVDNQHMLYHNY--TDNLKYSNYNKM------NDLSYNLHEK---------- 279

Qy        565 RSSCDNIMNK--KKNNLHLA---RHSVGSKLTMFSYDPQKN------------KDNTFFK 607
               ::|  | :|    : | :  |     : :|   |   :   : :|:                |||
Db        280 KNSFSNFINSGPRDNPMELCKKLKKAVEYKERVIDINKEKDFVLLGISKTCVKKCNTCSG 339

Qy        608 SNINK-----MEDN-----------TPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKD 651
               |: |       :||             ||||:   |  | | |            | : |:
Db        340 DNVTKDIDKCVEDEEKSKEGVILNYMKKDILF--YNTFNRNNN-----------DPNRKE 386

Qy        652 ENHNDNKINNGVINII---------NNSNVNSINNSNMNSINNSNMNSNSIY-------- 694
               :    :| |    ::::             :|:    :|: : ::    ::: | :|
Db        387 KPKECDKYNKDDVHVLCDHDHFSHSKSSHTTKNSNTKLYNVKEKHIHINKVYNNVYFVEG 446

Qy        695 -----------------KSNYNSNQSISDVQIRYVNEMDTSNKN---------------- 721
                                 :::|  : ::  :    | |:|   |
Db        447 QEKLYSPSIKEETQFYIQNDYKHDDNVKMLSYNYYNDMVYKNSKGMIDSLSTQHAFKGEE 506

Qy        722 ---------------NDNIFFDAI------------------------SCDNNMYPNI 740
                               |  :: |::                            |  |   |
Db        507 TVININKLRRRFSIMNRKVYSDSVLYFYGAPWWLNKIRRGQKIGQEKKHKKKDENKKKNK 566

Qy        741 TNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYN-------SHNLFQPDNNKNTS 793
               | | ||||:|| || :    |      |     |:      |     |   :| : : : |:|
Db        567 KNKNKNNNNSNNINNKHGRVIQYTDEKIQND----YCKNKESSKRGNHKMMRKEKNLNSS 622

Qy        794 NI-----------QNIN-TNKNNQDGNVNYSMNFCHYN--------------------LN 821
               :             :| | | |   :|         :|: |                    ||
Db        623 LLSINGKCYNKWKKNYNKTRKPKNEGRKGEKYIYCYENIKILEDIKDRFFNDHKRNNILN 682

Qy        822 DKNYL------------IDLNNKEQ-----------------KDKNIHGCDN-------- 844
               ::|::             ::  |||:                   ||    ||
Db        683 EENFIKEHQINGRNKEHVNEKNKEEDTFNISKENTKEGSYIITHKNKRNMDNIKIGRYDN 742

Qy        845 -------------------------------------------NIIQNRNDFEKKKKTN-- 860
                                                          |:|:  :  |  |  :||
Db        743 INDKKEFSSNILYKCVKKNDKINKSQTSLFFEFMKGKGDQKHNVIKKEDVFIKTFRTNKS 802

Qy        861 ----------------------------FYN-------------NNNIVI---------- 869
                                          ||                |:||
Db        803 PTELTKKISDYKCNLLYTSLDRIHKNVSIYNERIERTKHVPQKKNDNIDIRGIYKSYNFF 862

Qy        870 VNNNMGNNNSPRMKYGLCGSHTSIDNMKN--------------NEMKNNEMKDNEMKD- 913
               : || |: |        |  ::: | |||                 :: |    :|  | |:
Db        863 KSMNMMNSLSKCYHTKTC-DYSNYDFMKNKMSKKAQNKLVSKCISKYKKKAIKKKERKET 921

Qy        914 -----------NHI-----------------KSNNNNSSSSSSSNNNIYNNI-NDDDTFQ 944
                         ||                  | || :  |: :| :   |  |  ::   :
Db        922 TTTKKKYIYRKNEISISFDGNVFGHENRKRTKENNKSKESAYTSKSRKNNKIKGEEKKTK 981

Qy        945 NDYCHNDNTFTIRR------KNNTNINSNIYQ-----NDDIIYTINSLNDYMSNTLLHFK 993
               |    :: :|:       :|  :|  |:  |              :|  || ||  ||  |:
Db        982 RSLC----SYKLRKMKHLCVENKMHIKKNVRQIIKKKKKNIYKTIKCLNSY--KTLIDQV 1035

Qy        994 EKYTYPTLSTNEDIYNKEMEGKH-IRLDDQDKYDDNDN-NNVDNNNKNNVDNNVDNNNVD 1051
                :  :  ||:  :  |:  |    |  :  |  ::|  |  || ||||  ||  |||  || |  |
Db       1036 NVKGDEEHKLSNHVNNKKKKKKNCINENNDDNNNDNYNDNNDDNNNDNYNDNNNDNYN-D 1094

Qy       1052 NNVDNNDKNNVDNNNVDNDD-----------DDVDFHNIKNFNNNEYLSYFQKNVDTII 1099
               || |:||  || ||||   |:|              ::::   |::       |     :||  |
Db       1095 NNNDHNDHNNNDNNNDLNNDHRNDNNQREHSCEEINIQNVEQKCEGEKYEGKRKNKYTYY 1154

Qy       1100 NNCLNSLDISSMYDDTK------------EILNNILLSKYKAEKDNVIKK----YINE- 1141
               ||          : ::||              | :|    :   | ||     :||    |: |
Db       1155 NNYYKINGKNEIHDDYNIKSHGSRINYNIFNIKDNKHNNNDKGEKSCELKKCSIPYVKEK 1214

Qy       1142 -DIKNMSLEEI------DKTAQSIYE------KR----KVLLTKLLLLFKKNVDTQINNE 1184
               :::|  :  |  |       ||:       :|:        ||      ||:|  :     :|  |
Db       1215 YNLENNTYEIIGLIYYGDKS--QVYKCINMNNKRVYAMKVVLKECNEIFVDNF------- 1265

Qy       1185 TSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVE 1244
               ::|  |    :          |||                 |:|  :    |:|  :     |       ::::
Db       1266 ---IKKYLFL----KNNPHK---------------NIISIYDIFCNNN-----YICIIMD 1298

Qy       1245 YQNKSANSVLLN--VSSNNGDIILLNKKLVQDNIKNSMD--HNN--IHK 1287
               |     | ||:  :|      | :  :   |:   ||   ::||  |:|   ||:
Db       1299 Y---CEGSTLLDYFMSLVPGSLDVYEIKKIMKNIFIALDFFHSNNIIHR 1344
```

```
RESULT 39
E71619
RAD2 endonuclease PFB0265c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: E71619
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: E71619
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1516 <GAR>
A;Cross-references: GB:AE001383; GB:AE001362; NID:g3845135; PIDN:AAC71842.1; PID:g384513
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0265c

  Query Match             7.8%;  Score 534.5;  DB 2;  Length 1516;
  Best Local Similarity   21.6%;  Pred. No. 3.8e-15;
  Matches  326;  Conservative  225;  Mismatches  488;  Indels  469;  Gaps   77;

Qy        137 DNTKGIFISFYEI-YCGKLYDLLQ-----------KRKMVAALENGKKEVVVKDLKILR 183
Db         76 DNIKKAHLYFFFLRICKLLYYNIRPIFIFDGNPPELKRKTIFQRNIKKRNYEEKFKK--- 132

Qy        184 VLTKEELILKMIDGVLLRKI-GVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAG 242
Db        133 --TAEKLVYNYYQRTLLNSMKSKNKKNDNSN------NIEDKTNTPNKT---------N 174

Qy        243 SERGADTVSQNKQTQTDGANINR-SLLALKECIRAMDSDKN-------HIPFRDSELTKV 294
Db        175 TQNKSNTQNKSNTPNKINADISKSSLIQIYDDIKEKDKSLNSLVEHVGNVPVSVKDVLTI 234

Qy        295 LR-DIFVGKSKSIMIANISPTISCCEQ------TLNTLRYSSRVKNFKNKSTCI------ 341
Db        235 CNDDLSKIKNKIFMITDFGPVLFLGEQDGDMGTVENINKLDNRNKDENNLSYSINYNKVQ 294

Qy        342 ----NEEDDTNTERISI------------------------------LDSKGSEMNASS 366
Db        295 DVNNNNDDDKDKDKENINEVRRDQKNYVYKNKENINNIYLDDDDEKEDIQNKNGVYNNDD 354

Qy        367 IENVVIKSNHL--------------------------LSNNNNN--------KINRGKI 391
Db        355 IDEQIIRKKHMARKKYYESIPKTFKGFLCMRRPVDIIDISNYNTEMLEISETLKVHENKF 414

Qy        392 NDKI----ERN-------NILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINY 440
Db        415 KQHLNVLDENNSTPVVNMNLLKNINYKK---------------ND-DLIEGGEKKSFINL 458

Qy        441 ----------KSTLYNDNTINK-KHN----NNNNNNNDNNDNNNDNNNNNNNDSSSMVNN 485
Db        459 INVDSCYSSSNSRLENDENIERGKINMFITNDEKSININNYNNNNNNNNNNNDNNMDNN 518

Qy        486 --MINHMINN-NINN---NINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMN 539
Db        519 DVIIEHNKNNMNIYDNKYNVECSSEKINDNGISNKNINILELPN----------NLDTSN 568

Qy        540 CHLNNND--KSIFLHKKNLRDNI-----------------------KLKNRSSCDNIMN-- 573
Db        569 IFLEGKDEYKVYYVNKEEIRIPLFKEINKEIFEKLPLKLQYQILQDIKEEWYTDNRIKAI 628

Qy        574 KKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNG 633
Db        629 KSKDDM-----DVFSQVQLETYVRMIKTDFEIEKLKIKMAE-----------NIQSVEG 671

Qy        634 NVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSI-NNSNMNSNS 692
Db        672 ELLI-----------NKDLSKNTDNINIKDYNVLQKKKSKKKKKKFLNDILNTYNFTTES 720

Qy        693 IYKSNY-NSNQSISDV--QIRYVNEMDTSNKNNDNI--------FFDAISCDNNMYPNIT 741
Db        721 KYQDLYVKGEESKEDIKNQIDFVTQ--ECYRNNDIIRDTHDKSDIFKNIKIDNNKKYEIY 778

Qy        742 N--------NNNNNNNNNNNNNN---IDVENYNNR-----------DGTNNSMKLYAYN 779
Db        779 NLELEQEEINEKKNYNKNNNDSNKTFFLKIENEFKKDLLLDDSQIFGDSLLADIKEYNYT 838

Qy        780 SHNLFQPDNNKNTS------NIQNIN---TNKNNQDGNVNYSMNFCHYNLNDKNYLIDLN 830
Db        839 ADNL--DNNNENKSLYEDGENFITRNEPITNEYEEKNNIIYISDEQKYNEED---IIFKD 893

Qy        831 NKEQKDKN-------IHGCD-------NNIIQ---NRND--------------FEKKKK 858
Db        894 KIKEKEKNNDTSSDDFENCSVQEKIYVNEKIEEYNNKNDDKSSSSSSIILEEIKYKKEKK 953

Qy        859 TNFYNNNNIVIV-----NNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNN--EMKDNEM 911
Db        954 DELVSPNLCVLLDEFEHSNDLENNY----------ISVSSDDMKTNVSKNNITGVKENKV 1003

Qy        912 KDNHI---------------KSNNNNSSSSSSNNNIYNNINDDDTFQN---DYC----- 948
Db       1004 DKTNVEYDKKGDDGVIEISFEDSHKLEESKFDDNNNIYD--NDDELEKNLSKDYISDVDK 1061

Qy        949 -HNDNTFTIRR----KNNTNINSNI-----YQNDDIIYTINS--LNDYMSNTLLHFKEKY 996
Db       1062 NHVNNIYNIERGEDERENEFVENKIQSTESHKSNEFICTENKSLRKQYMS------KEDI 1115

Qy        997 TYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNN----KNNVDNNV------D 1046
Db       1116 SNVRILKSDDINNLSKQNYFEILLDKKQVMDNFQMNIEQNNDKLKEDKLDEGAYFEYLED 1175

Qy       1047 NNNVDNNVDNNDKNN---------VDNNNVD-NDDDDVDFHNIKNFNNNEYL-------- 1088
Db       1176 NKIIDSYIKETNKENEELIKEYKKLKKNNIEINDEMNDDIKLLLNFFGIPYIQSPCEAEA 1235

Qy       1089 --SYF-QKN-VDTIINNCLNSLDISSMYDDTKEILNNILLSKYKA---EKDNVIKK--YI 1139
Db       1236 QCSYLNNKNYCDAIISDDSDVLVFSG-----KTVIKNFFNKKKTVEVYEKKAIEEKLGLY 1290

Qy       1140 NEDIKNMSL------------------EEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQI 1181
Db       1291 QEELINISLLCGCDYTIGVHGIGIVNALEIIKAFPN-FEDLKILKDIVSNPFRK-IDKNM 1348

Qy       1182 NNET----SDLRKDLVMCHICNNN-PD-DQFHFYAYSRLEKDIINLIMLRQIWCESENLR 1235
Db       1349 YNEEIQQFLNTHKNYKLNWIFPNNFPDREVYKCFKYPKVCTDI------KKFEWHVPDIK 1402

Qy       1236 LLYQFL----------VVEYQNKSANSVLLNVSSNNGDI-----ILLNKKLVQDNIKNSM 1280
Db       1403 SITKFLHKTTNISEEKVLNVLNPILQKYNVNVRTYQSKIEDFFPLLEKKRKTVDDLIDHI 1462

Qy       1281 DHNNIHKK 1288
Db       1463 RANNKQKR 1470

RESULT 40
T47525
kinesin-related protein-like - Arabidopsis thaliana
N;Alternate names: protein F16L2.60
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C;Accession: T47525
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.;
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24468
A;Accession: T47525
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1058 <JOR>
A;Cross-references: EMBL:AL162459
A;Experimental source: cultivar Columbia; BAC clone F16L2
```

```
C;Genetics:
A;Map position: 3
A;Introns: 56/2; 99/3; 151/3; 310/3; 343/2; 395/3; 419/1; 444/3; 462/3; 489/3; 527/1; 55
A;Note: F16L2.60
C;Superfamily: kinesin-related protein KIP1; kinesin motor domain homology

  Query Match             7.8%;  Score 534;  DB 2;  Length 1058;
  Best Local Similarity   22.4%;  Pred. No. 2.7e-15;
  Matches  227;  Conservative  192;  Mismatches  412;  Indels  184;  Gaps   36;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMT-----KYIERHEFIVDK 59
              ::|::| ||||| | :       :|:    |           | :|  |     |:|:|| |  ||
Db         49 VQVILRCRPLSEDEARIHTPVVISCNEN--------RREVAATQSIAGKHIDRH-FAFDK 99

Qy         60 VFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTP-- 117
              ||          :|:  | |:: :: | |   |: ||||||:||||| |          : |
Db        100 VFGPASQQKDLYDQAICPIVFEVLE-GYNCTIFAYGQTGTGKTYTMEGGARKKNGEFPSD 158

Qy        118 -GIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLL-----------QKRKMVA 165
               |:   |   || |       : :   : ::| |:| :: |||              : :| :|
Db        159 AGVIPRAVKQIFDILEAQGAEYS--MKVTFLELYNEEISDLLAPEETIKFVDEKSKKSIA 216

Qy        166 ALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKD 225
               :|:|| | |: |:   | |  |:   :  |   |:       | :|||||:| :| :
Db        217 LMEDGKGSVFVRGLEEEIVSTANEIYKILEKGSAKRRTAETLLNKQSSRSHSIFSITI-H 275

Qy        226 INKNT-------SLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMD 278
              | :||        ||:  :|||||  : : ::  :  :  |  ||:||| |    | |:
Db        276 IKENTPEGEEMIKCGKLNLVDLAGSENISRSGAREGRAREAG-EINKSLLTLGRVINALV 334

Qy        279 SDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKS 338
                 |||:|||:||:|::||:     ||:|:  :|| |||:|  | |:|:|| |: | || |||
Db        335 EHSGHIPYRDSKLTRLLRESLGGKTKTCVIATISPSIHCLEETLSTLDYAHRAKNIKNKP 394

Qy        339 TCINE---EDDTNTERISILDSKGSEMNASSIEN-VVIKSNHLLSNNNNNKINRGKINDK 394
                ||:   :      :  | :|    |: |:  :| : |   :   :     |     : :|
Db        395 E-INQKMMKSAVMKDLYSEIDRLKQEVYAAREKNGIYIPKDRYIQEEAEKK----AMAEK 449

Qy        395 IERNNILKNKSFDKPREGFTSTFGKYSSLND--IDKIKKNKKKGLINYKSTLYNDNTINK 452
              |||   |:::| ||          :       |       :|::| :|| |   : :|::    : :
Db        450 IERLE-LQSESKDKRVVDLQELYNSQQILTAELSEKLEKTEKK-LEETEHSLFD---LEE 504

Qy        453 KHNNNNNNNNDNNDNNNDNNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNN 512
              |:  |     :                            :|:| : :  ::
Db        505 KYRQANATIKEK----------------------EFVISNLLKSEKSLVER-------- 533

Qy        513 NSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIM 572
                           ||   ::  |         ::| |     |    :|  |:  ||
Db        534 ------------AFQLRTELESA--------SSDVSNLFSKIERKDKIEDGNRFLIQKFQ 573

Qy        573 NKKKNNLHLARHSVGSKLT------------MFSYDPQKNKDNTFFKSNINKMEDNTPKD 620
              ::      | |    :| | :|          | |:    |::      :   ::|::
Db        574 SQLTQQLELLHKTVASSVTQQEVQLKHMEEDMESFVSTKSEATEELRDRLSKLKR----- 628

Qy        621 ILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINN--SNVNSINNS 678
               :| |   :  :: |: : |: |:   |:   |   :| :: : |:      |   : :
Db        629 -VYGS-GIEALD-NIAVKLDGNSQSTFSSL--NSEVSKHSHELENVFKGFASEADMLLQD 683

Qy        679 NMNSINNSN-------MNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAIS 731
                :|:|              :     :: :|:| | : :   :||           | :| :
Db        684 LQSSLNKQEEKLITFAQQQRKAHSRAVDTARSVSKVTVEFFKTLDTHATKLTGIVEEAQT 743

Qy        732 CDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKN 791
                ::             |      |     |    | |  :   ::    |   |:| :  : :
Db        744 VNHKKLSEFENKFEECAANEERQLLEKVAELLANSNARKKNLVQMAV--HDLRESASTRT 801

Qy        792 TSNIQNINTNKNNQDG-NVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIH----GCDNNI 846
              |:    ::| :::        :|:    |      : ::  | :   |     | :         |
Db        802 TTLQHEMSTMQDSTSSIKAEWSI---HMEKTESSHHEDTSAVESGKKAMQEVLLNCLEKT 858

Qy        847 IQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEM 906
                : : : | ::       : : :   ||:                   |:|::    |  ||
Db        859 EMSAHQWRKAQE------SLVSLERNNV----------------ASVDSIVRGGMDANE- 895

Qy        907 KDNEMKDNHIKSNNNNSSSSS-----SSNNNIYNNINDDDTFQNDYCHNDNTFTI 956
                       :::|   : :  |||       ::|:::  :|:          || |    |:  |
Db        896 --------NLRSQFSTAVSSSLDVFDAANSSLLTSIDHSLQLDNDACTKVNSMII 942

Search completed: October  2, 2003, 16:21:28
Job time : 94 secs
```

OM protein - protein search, using sw model

Run on: October 2, 2003, 16:17:11 ; Search time 33 Seconds
(without alignments)
1651.406 Million cell updates/sec

Title: US-10-006-780-2
Perfect score: 6842
Sequence: 1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	768.5	11.2	363	4	US-09-967-908A-10	Sequence 10, Appl
2	768.5	11.2	381	4	US-09-967-908A-6	Sequence 6, Appli
3	768.5	11.2	1368	4	US-09-967-908A-2	Sequence 2, Appli
4	758.5	11.1	343	4	US-09-967-908A-8	Sequence 8, Appli
5	758.5	11.1	373	4	US-09-967-908A-4	Sequence 4, Appli
6	758.5	11.1	730	4	US-09-291-170A-3	Sequence 3, Appli
7	758.5	11.1	730	4	US-09-724-884-3	Sequence 3, Appli
8	752.5	11.0	405	4	US-09-594-669-12	Sequence 12, Appl
9	752.5	11.0	434	4	US-09-594-669-10	Sequence 10, Appl
10	752.5	11.0	473	4	US-09-594-669-8	Sequence 8, Appli
11	752.5	11.0	723	4	US-09-594-669-14	Sequence 14, Appl
12	750.5	11.0	725	4	US-09-594-669-16	Sequence 16, Appl
13	750.5	11.0	725	4	US-09-595-684B-33	Sequence 33, Appl
14	745.5	10.9	362	4	US-09-594-669-6	Sequence 6, Appli
15	745.5	10.9	391	4	US-09-594-669-4	Sequence 4, Appli
16	745.5	10.9	430	4	US-09-594-669-2	Sequence 2, Appli
17	736	10.8	678	4	US-09-595-684B-25	Sequence 25, Appl
18	696.5	10.2	319	4	US-09-724-510-2	Sequence 2, Appli
19	696.5	10.2	319	4	US-09-723-216-2	Sequence 2, Appli
20	696.5	10.2	319	4	US-09-675-227-2	Sequence 2, Appli
21	589.5	8.6	303	4	US-09-722-129-4	Sequence 4, Appli
22	589.5	8.6	492	4	US-09-722-129-2	Sequence 2, Appli
23	561.5	8.2	1111	4	US-09-914-259-28	Sequence 28, Appl
24	538.5	7.9	1073	3	US-09-541-782-6	Sequence 6, Appli
25	538.5	7.9	1073	4	US-09-723-820-6	Sequence 6, Appli
26	538.5	7.9	2184	4	US-09-417-485D-6	Sequence 6, Appli
27	533.5	7.8	2391	2	US-08-446-855A-2	Sequence 2, Appli
28	533.5	7.8	2391	3	US-09-150-741-2	Sequence 2, Appli
29	500.5	7.3	967	4	US-09-914-259-21	Sequence 21, Appl
30	500	7.3	1184	3	US-09-541-782-2	Sequence 2, Appli
31	500	7.3	1184	4	US-09-723-820-2	Sequence 2, Appli
32	482	7.0	2662	4	US-09-595-684B-31	Sequence 31, Appl
33	465.5	6.8	367	4	US-09-724-511-2	Sequence 2, Appli
34	465.5	6.8	367	4	US-09-723-097-2	Sequence 2, Appli
35	465.5	6.8	367	4	US-09-632-344-2	Sequence 2, Appli
36	465	6.8	1038	3	US-09-541-782-4	Sequence 4, Appli
37	465	6.8	1038	4	US-09-723-820-4	Sequence 4, Appli
38	464.5	6.8	355	4	US-09-724-511-4	Sequence 4, Appli
39	464.5	6.8	355	4	US-09-723-097-4	Sequence 4, Appli
40	464.5	6.8	355	4	US-09-632-344-4	Sequence 4, Appli
41	460.5	6.7	1388	4	US-09-572-191-2	Sequence 2, Appli
42	460.5	6.7	1388	4	US-09-723-262-2	Sequence 2, Appli
43	460.5	6.7	1388	4	US-09-723-219-2	Sequence 2, Appli
44	459	6.7	1066	3	US-09-541-782-8	Sequence 8, Appli
45	459	6.7	1066	4	US-09-723-820-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-967-908A-10
; Sequence 10, Application US/09967908A
; Patent No. 6514738
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christopher
; APPLICANT: Guo, Jun
; APPLICANT: Freedman, Richard
; APPLICANT: Patel, Umesh A.
; APPLICANT: Davies, Katherine A.
; TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
; FILE REFERENCE: 020552-002400US
; CURRENT APPLICATION NUMBER: US/09/967,908A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/675,227
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Peptide
; OTHER INFORMATION: fragment E213 - S566.
US-09-967-908A-10

Query Match 11.2%; Score 768.5; DB 4; Length 363;
Best Local Similarity 47.5%; Pred. No. 6.4e-49;
Matches 164; Conservative 58; Mismatches 116; Indels 7; Gaps 3;

```
Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| ||||||    | :: : :||||::  || : | :   ||:|:|| :| |  |:|| :
Db         12 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 71

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                 |  ||   |  |||   :: ||  :|||||||:||||||:|:      : ||:: |
Db         72 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 125

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
               | |||  |        ::||||||||:||||| :|| : | |: |  | :  |: |:
Db        126 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 185

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
               | : | |:  :: |   |   |    | :||||||:: | :||  |  | |:|:|||||||
Db        186 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 244

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
               || ||     ::||: :|| ||:|||||||||:| :  | ||| |:||:||:| |:| :
```

```
Db        245 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 304

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
              |: |||||||:     | ||||||: |||  |      |       |
Db        305 KTCMIANISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRN 349
```

RESULT 2
US-09-967-908A-6
; Sequence 6, Application US/09967908A
; Patent No. 6514738
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christopher
; APPLICANT: Guo, Jun
; APPLICANT: Freedman, Richard
; APPLICANT: Patel, Umesh A.
; APPLICANT: Davies, Katherine A.
; TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
; FILE REFERENCE: 020552-002400US
; CURRENT APPLICATION NUMBER: US/09/967,908A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/675,227
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 381
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Peptide
;   OTHER INFORMATION: fragment V195-S566 with flanking vector sequence.
US-09-967-908A-6

```
  Query Match            11.2%;  Score 768.5;  DB 4;  Length 381;
  Best Local Similarity   47.5%;  Pred. No. 6.8e-49;
  Matches  164;  Conservative   58;  Mismatches  116;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| |||||     | :: : :||||::  || : | :   ||:|:|| :| |  |:|| :
Db         30 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 89

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                 | ||  |  |||  :: ||    :|||||||||:|||||||:|:       : ||:: |
Db         90 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 143

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              | ||| | :            ::|||||||||:||||| :|| : | |: |  | :  |: |:
Db        144 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 203

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              | : | |:  :: |    |   |     | :||||||:: | :|| | |  |:|:|||||||
Db        204 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 262

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              || ||     ::||: :|| ||:|||||||||||:| :   | ||| |:||:||:| |:| :
Db        263 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 322

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
              |: |||||||:     | ||||||: |||  |      |       |
Db        323 KTCMIANISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRN 367
```

RESULT 3
US-09-967-908A-2
; Sequence 2, Application US/09967908A
; Patent No. 6514738
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christopher
; APPLICANT: Guo, Jun
; APPLICANT: Freedman, Richard
; APPLICANT: Patel, Umesh A.
; APPLICANT: Davies, Katherine A.
; TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
; FILE REFERENCE: 020552-002400US
; CURRENT APPLICATION NUMBER: US/09/967,908A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/675,227
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1368
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: KinI-3 motor
;   OTHER INFORMATION: domain fragment
US-09-967-908A-2

```
  Query Match            11.2%;  Score 768.5;  DB 4;  Length 1368;
  Best Local Similarity   47.5%;  Pred. No. 4.1e-48;
  Matches  164;  Conservative   58;  Mismatches  116;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| |||||     | :: : :||||::  || : | :   ||:|:|| :| |  |:|| :
Db        223 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 282

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                 | ||  |  |||  :: ||    :|||||||||:|||||||:|:       : ||:: |
Db        283 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 336

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              | ||| | :            ::|||||||||:||||| :|| : | |: |  | :  |: |:
Db        337 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 396

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              | : | |:  :: |    |   |     | :||||||:: | :|| | |  |:|:|||||||
Db        397 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 455

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              || ||     ::||: :|| ||:|||||||||||:| :   | ||| |:||:||:| |:| :
Db        456 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 515

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
              |: |||||||:     | ||||||: |||  |      |       |
Db        516 KTCMIANISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRN 560
```

RESULT 4
US-09-967-908A-8
; Sequence 8, Application US/09967908A
; Patent No. 6514738
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christopher
; APPLICANT: Guo, Jun
; APPLICANT: Freedman, Richard
; APPLICANT: Patel, Umesh A.
; APPLICANT: Davies, Katherine A.
; TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
; FILE REFERENCE: 020552-002400US
; CURRENT APPLICATION NUMBER: US/09/967,908A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/675,227
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 343
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Peptide

```
;   OTHER INFORMATION: fragment E213-S546.
US-09-967-908A-8

  Query Match             11.1%;  Score 758.5;  DB 4;  Length 343;
  Best Local Similarity   48.9%;  Pred. No. 3.2e-48;
  Matches  161;  Conservative   58;  Mismatches  103;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db         12 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 71

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
Db         72 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 125

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
Db        126 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 185

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
Db        186 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 244

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
Db        245 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 304

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVK 332
Db        305 KTCMIANISPSHVATEHTLNTLRYADRVK 333

RESULT 5
US-09-967-908A-4
; Sequence 4, Application US/09967908A
; Patent No. 6514738
; GENERAL INFORMATION:
;  APPLICANT: Cytokinetics, Inc.
;  APPLICANT:  Beraud, Christopher
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  APPLICANT:  Patel, Umesh A.
;  APPLICANT:  Davies, Katherine A.
;  TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
;  FILE REFERENCE: 020552-002400US
;  CURRENT APPLICATION NUMBER: US/09/967,908A
;  CURRENT FILING DATE:  2001-09-28
;  PRIOR APPLICATION NUMBER: US 09/675,227
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 373
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Fragment
;   OTHER INFORMATION: encoding residues D183-L546 with flanking vector
;   OTHER INFORMATION: sequence.
US-09-967-908A-4

  Query Match             11.1%;  Score 758.5;  DB 4;  Length 373;
  Best Local Similarity   48.9%;  Pred. No. 3.6e-48;
  Matches  161;  Conservative   58;  Mismatches  103;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db         42 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 101

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
Db        102 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 155

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
Db        156 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 215

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
Db        216 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 274

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
Db        275 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 334

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVK 332
Db        335 KTCMIANISPSHVATEHTLNTLRYADRVK 363

RESULT 6
US-09-291-170A-3
; Sequence 3, Application US/09291170A
; Patent No. 6410687
; GENERAL INFORMATION:
;  APPLICANT:  Vale, Ronald D.
;  APPLICANT:  Hartman, James J.
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION:  Assays for the Detection of Microtubule
;  TITLE OF INVENTION:  Depolymerization Inhibitors
;  FILE REFERENCE:  18557B-000510US
;  CURRENT APPLICATION NUMBER: US/09/291,170A
;  CURRENT FILING DATE:  1999-04-13
;  PRIOR APPLICATION NUMBER:  US 60/081,734
;  PRIOR FILING DATE:  1998-04-14
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 730
;   TYPE: PRT
;   ORGANISM: Xenopus laevis
;   FEATURE:
;   OTHER INFORMATION: Xenopus kinesin central motor 1 (XKCM1)
US-09-291-170A-3

  Query Match             11.1%;  Score 758.5;  DB 4;  Length 730;
  Best Local Similarity   42.1%;  Pred. No. 9.3e-48;
  Matches  175;  Conservative   84;  Mismatches  130;  Indels   27;  Gaps   12;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db        262 RICVCVRKRPLNKQELSKKEIDIISVPSKNIVLVHEPKLKVDLTKYLENQAFRFDFSFDE 321

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
Db        322 TATNEVVYRFTARPLVQSIFEGG-KATCFAYGQTGSGKTHTMGGDFSGKSQNVSKGVYAF 380

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
Db        381 ASRDVFLLL---DQPRYKHLDLDVFVTFFEIYNGKVFDLLNKKTKLRVLEDAKQEVQVVG 437

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSL-GKIA 236
Db        438 LLEKQVISADD-VFKMIEIGSACRTSGQTFANTSSSRSHACLQIILR---RGSKLHGKFS 493

Qy        237 FIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLR 296
Db        494 LVDLAGNERGVDTASADRITRMKGAEINRSLLALKECIRALGQNKSHTPFRESKLTQILR 553

Qy        297 DIFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISIL 355
Db        554 DSFIGENSRTCMIAMLSPGFNSCEYTLNTLRYADRVKELSPQNA------ETNDDNLQME 607

Qy        356 DSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
Db        608 DSGGSH---ASIEGLQLQDDFLLKDEELSTHN--SFQDALNRVGELEDKAVDELRE 658
```

```
RESULT 7
US-09-724-884-3
; Sequence 3, Application US/09724884
; Patent No. 6429304
; GENERAL INFORMATION:
;  APPLICANT: Vale, Ronald D.
;  APPLICANT:  Hartman, James J.
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: Assays for the Detection of Microtubule
;  TITLE OF INVENTION:  Depolymerization Inhibitors
;  FILE REFERENCE: 18557B-000510US
;  CURRENT APPLICATION NUMBER: US/09/724,884
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: 09/291,170
;  PRIOR FILING DATE: 1999-04-13
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 730
;   TYPE: PRT
;   ORGANISM: Xenopus laevis
;   FEATURE:
;   OTHER INFORMATION: Xenopus kinesin central motor 1 (XKCM1)
US-09-724-884-3

  Query Match             11.1%;  Score 758.5;  DB 4;  Length 730;
  Best Local Similarity   42.1%;  Pred. No. 9.3e-48;
  Matches  175;  Conservative   84;  Mismatches  130;  Indels   27;  Gaps   12;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db        262 RICVCVRKRPLNKQELSKKEIDIISVPSKNIVLVHEPKLKVDLTKYLENQAFRFDFSFDE 321

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
Db        322 TATNEVVYRFTARPLVQSIFEGG-KATCFAYGQTGSGKTHTMGGDFSGKSQNVSKGVYAF 380

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
Db        381 ASRDVFLLL---DQPRYKHLDLDVFVTFFEIYNGKVFDLLNKKTKLRVLEDAKQEVQVVG 437

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSL-GKIA 236
Db        438 LLEKQVISADD-VFKMIEIGSACRTSGQTFANTSSSRSHACLQIILR---RGSKLHGKFS 493

Qy        237 FIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLR 296
Db        494 LVDLAGNERGVDTASADRITRMKGAEINRSLLALKECIRALGQNKSHTPFRESKLTQILR 553

Qy        297 DIFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISIL 355
Db        554 DSFIGENSRTCMIAMLSPGFNSCEYTLNTLRYADRVKELSPQNA------ETNDDNLQME 607

Qy        356 DSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
Db        608 DSGGSH---ASIEGLQLQDDFLLKDEELSTHN--SFQDALNRVGELEDKAVDELRE 658

RESULT 8
US-09-594-669-12
; Sequence 12, Application US/09594669
; Patent No. 6331424
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Sakowicz, Roman
;  TITLE OF INVENTION:  No. 6331424el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1042
;  CURRENT APPLICATION NUMBER: US/09/594,669
;  CURRENT FILING DATE:  2000-06-15
;  PRIOR APPLICATION NUMBER:  US 09/295,612
;  PRIOR FILING DATE:  1999-04-20
;  PRIOR APPLICATION NUMBER:  US 09/314,464
;  PRIOR FILING DATE:  1999-05-18
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 405
;   TYPE: PRT
;   ORGANISM: Human
US-09-594-669-12

  Query Match             11.0%;  Score 752.5;  DB 4;  Length 405;
  Best Local Similarity   42.1%;  Pred. No. 1.1e-47;
  Matches  174;  Conservative   75;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db          3 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 62

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
Db         63 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 121

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
Db        122 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 180

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
Db        181 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 237

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
Db        238 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 297

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
Db        298 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 348

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
Db        349 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 399

RESULT 9
US-09-594-669-10
; Sequence 10, Application US/09594669
; Patent No. 6331424
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Sakowicz, Roman
;  TITLE OF INVENTION:  No. 6331424el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1042
;  CURRENT APPLICATION NUMBER: US/09/594,669
;  CURRENT FILING DATE:  2000-06-15
;  PRIOR APPLICATION NUMBER:  US 09/295,612
;  PRIOR FILING DATE:  1999-04-20
;  PRIOR APPLICATION NUMBER:  US 09/314,464
;  PRIOR FILING DATE:  1999-05-18
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 434
;   TYPE: PRT
;   ORGANISM: Human
US-09-594-669-10

  Query Match             11.0%;  Score 752.5;  DB 4;  Length 434;
  Best Local Similarity   42.1%;  Pred. No. 1.2e-47;
  Matches  174;  Conservative   75;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
```

```
                 :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|   |  |  ||:
Db           32 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 91

Qy           64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
                 |  |  ||  | :||:  ::| |   :|||||||||||||:|| |        |: : ||:
Db           92 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 150

Qy          123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
                |: |:|   |   | |   : ::::|:||| |||:||| |:  :  ||:||::| |  |:
Db          151 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 209

Qy          181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                    | : :: ::|||| |   |  |    |  |||||   | |:   |    || : :|
Db          210 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 266

Qy          240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
                |||:|||||| | ::||: :|| ||:||||||||||||:  :| | |||:|:||:|||| |
Db          267 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 326

Qy          300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
                :|: |:: ||| ||| || || |||||||: |||       |     |:          :: :
Db          327 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 377

Qy          359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                 || | |  |  :  :|               |: : :   |: |: :: :|
Db          378 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 428
```

RESULT 10
US-09-594-669-8
; Sequence 8, Application US/09594669
; Patent No. 6331424
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: No. 6331424el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1042
; CURRENT APPLICATION NUMBER: US/09/594,669
; CURRENT FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/295,612
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: US 09/314,464
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 473
;   TYPE: PRT
;   ORGANISM: Human
US-09-594-669-8

```
  Query Match             11.0%;  Score 752.5;  DB 4;  Length 473;
  Best Local Similarity   42.1%;  Pred. No. 1.4e-47;
  Matches  174;  Conservative   75;  Mismatches  143;  Indels   21;  Gaps   10;

Qy            4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                 :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|   |  |  ||:
Db           71 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 130

Qy           64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
                 |  |  ||  | :||:  ::| |   :|||||||||||||:|| |        |: : ||:
Db          131 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 189

Qy          123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
                |: |:|   |   | |   : ::::|:||| |||:||| |:  :  ||:||::| |  |:
Db          190 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 248

Qy          181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                    | : :: ::|||| |   |  |    |  |||||   | |:   |    || : :|
Db          249 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 305

Qy          240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
                |||:|||||| | ::||: :|| ||:||||||||||||:  :| | |||:|:||:|||| |
Db          306 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 365

Qy          300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
                :|: |:: ||| ||| || || |||||||: |||       |     |:          :: :
Db          366 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 416

Qy          359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                 || | |  |  :  :|               |: : :   |: |: :: :|
Db          417 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 467
```

RESULT 11
US-09-594-669-14
; Sequence 14, Application US/09594669
; Patent No. 6331424
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: No. 6331424el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1042
; CURRENT APPLICATION NUMBER: US/09/594,669
; CURRENT FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/295,612
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: US 09/314,464
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 723
;   TYPE: PRT
;   ORGANISM: Human
US-09-594-669-14

```
  Query Match             11.0%;  Score 752.5;  DB 4;  Length 723;
  Best Local Similarity   42.1%;  Pred. No. 2.6e-47;
  Matches  174;  Conservative   75;  Mismatches  143;  Indels   21;  Gaps   10;

Qy            4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                 :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|   |  |  ||:
Db          256 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 315

Qy           64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
                 |  |  ||  | :||:  ::| |   :|||||||||||||:|| |        |: : ||:
Db          316 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 374

Qy          123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
                |: |:|   |   | |   : ::::|:||| |||:||| |:  :  ||:||::| |  |:
Db          375 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 433

Qy          181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                    | : :: ::|||| |   |  |    |  |||||   | |:   |    || : :|
Db          434 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 490

Qy          240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
                |||:|||||| | ::||: :|| ||:||||||||||||:  :| | |||:|:||:|||| |
Db          491 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 550

Qy          300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
                :|: |:: ||| ||| || || |||||||: |||       |     |:          :: :
Db          551 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 601

Qy          359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                 || | |  |  :  :|               |: : :   |: |: :: :|
Db          602 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 652
```

RESULT 12
US-09-594-669-16
; Sequence 16, Application US/09594669
; Patent No. 6331424

```
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Sakowicz, Roman
;  TITLE OF INVENTION:  No. 6331424el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1042
;  CURRENT APPLICATION NUMBER: US/09/594,669
;  CURRENT FILING DATE:  2000-06-15
;  PRIOR APPLICATION NUMBER:  US 09/295,612
;  PRIOR FILING DATE:  1999-04-20
;  PRIOR APPLICATION NUMBER:  US 09/314,464
;  PRIOR FILING DATE:  1999-05-18
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 725
;   TYPE: PRT
;   ORGANISM: Human
US-09-594-669-16

  Query Match             11.0%;  Score 750.5;  DB 4;  Length 725;
  Best Local Similarity   41.9%;  Pred. No. 3.6e-47;
  Matches  173;  Conservative   76;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |   ||: |:|:: : | | : ||: |||:|||:|    |   | ||:
Db        258 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 317

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |   |  ||   | :||:  ::| |    :||||||||||||:|| |          |: : ||:
Db        318 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 376

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |     | |     : ::::|:|||  |||:||| |:   :    ||:||::| |   |:
Db        377 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 435

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                  | : :: ::||:|  |     |     |      |     ||||||    | |:     |      || : :|
Db        436 EHLVNSADD-VIKMLDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 492

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||  |  ::|:|: :||  ||:||||||||||||:    :|  |  |||:|:||:|:|||| |
Db        493 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 552

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||       |       |:                ::   :
Db        553 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 603

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                  || |  |   |   :    :|                    |: : :     |: |: :: :|
Db        604 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 654

RESULT 13
US-09-595-684B-33
; Sequence 33, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Ohashi, Cara
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  VAisberg, Eugeni
;  APPLICANT:  Wood, Kenneth
;  APPLICANT:  Yu, Ming
;  TITLE OF INVENTION: Human kinesins and methods of producing
;  TITLE OF INVENTION:  and purifying human kinesins
;  FILE REFERENCE: cytop036
;  CURRENT APPLICATION NUMBER: US/09/595,684B
;  CURRENT FILING DATE:  2002-06-24
;  PRIOR APPLICATION NUMBER: 09/295,612
;  PRIOR FILING DATE: 2000-04-20
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 33
;   LENGTH: 725
;   TYPE: PRT
;   ORGANISM: Human
US-09-595-684B-33

  Query Match             11.0%;  Score 750.5;  DB 4;  Length 725;
  Best Local Similarity   41.9%;  Pred. No. 3.6e-47;
  Matches  173;  Conservative   76;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |   ||: |:|:: : | | : ||: |||:|||:|    |   | ||:
Db        258 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 317

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |   |  ||   | :||:  ::| |    :||||||||||||:|| |          |: : ||:
Db        318 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 376

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |     | |     : ::::|:|||  |||:||| |:   :    ||:||::| |   |:
Db        377 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 435

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                  | : :: ::||:|  |     |     |      |     ||||||    | |:     |      || : :|
Db        436 EHLVNSADD-VIKMLDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 492

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||  |  ::|:|: :||  ||:||||||||||||:    :|  |  |||:|:||:|:|||| |
Db        493 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 552

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||       |       |:                ::   :
Db        553 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 603

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                  || |  |   |   :    :|                    |: : :     |: |: :: :|
Db        604 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 654

RESULT 14
US-09-594-669-6
; Sequence 6, Application US/09594669
; Patent No. 6331424
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Sakowicz, Roman
;  TITLE OF INVENTION:  No. 6331424el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1042
;  CURRENT APPLICATION NUMBER: US/09/594,669
;  CURRENT FILING DATE:  2000-06-15
;  PRIOR APPLICATION NUMBER:  US 09/295,612
;  PRIOR FILING DATE:  1999-04-20
;  PRIOR APPLICATION NUMBER:  US 09/314,464
;  PRIOR FILING DATE:  1999-05-18
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 362
;   TYPE: PRT
;   ORGANISM: Human
US-09-594-669-6

  Query Match             10.9%;  Score 745.5;  DB 4;  Length 362;
  Best Local Similarity   45.7%;  Pred. No. 3.2e-47;
  Matches  168;  Conservative   65;  Mismatches  116;  Indels   19;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |   ||: |:|:: : | | : ||: |||:|||:|    |   | ||:
Db          3 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 62

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |   |  ||   | :||:  ::| |    :||||||||||||:|| |          |: : ||:
```

```
Db         63 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 121

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |    | |    : ::::|:||| |||:||| |:  :   ||:||::| |  |:
Db        122 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 180

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                 | : :: ::||||  |    |    |     |  ||||||   | |:    |     || : :|
Db        181 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 237

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:|||||| | ::||: :|| ||:|||||||||||:  :| | |||:|:||:|||| |
Db        238 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 297

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||        |      |:           :: :
Db        298 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 348

Qy        359 GSEMNASS 366
                 || | |
Db        349 TEEMEACS 356
```

RESULT 15
US-09-594-669-4
; Sequence 4, Application US/09594669
; Patent No. 6331424
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: No. 6331424el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1042
; CURRENT APPLICATION NUMBER: US/09/594,669
; CURRENT FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/295,612
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: US 09/314,464
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Human
US-09-594-669-4

```
  Query Match             10.9%;  Score 745.5;  DB 4;  Length 391;
  Best Local Similarity   45.7%;  Pred. No. 3.5e-47;
  Matches  168;  Conservative   65;  Mismatches  116;  Indels   19;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db         32 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 91

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |  |  ||  | :||:  ::| |    :||||||||||||:|| |      |: : ||:
Db         92 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 150

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |    | |    : ::::|:||| |||:||| |:  :   ||:||::| |  |:
Db        151 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 209

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                 | : :: ::||||  |    |    |     |  ||||||   | |:    |     || : :|
Db        210 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 266

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:|||||| | ::||: :|| ||:|||||||||||:  :| | |||:|:||:|||| |
Db        267 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 326

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||        |      |:           :: :
Db        327 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 377

Qy        359 GSEMNASS 366
                 || | |
Db        378 TEEMEACS 385
```

RESULT 16
US-09-594-669-2
; Sequence 2, Application US/09594669
; Patent No. 6331424
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: No. 6331424el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1042
; CURRENT APPLICATION NUMBER: US/09/594,669
; CURRENT FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 09/295,612
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: US 09/314,464
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Human
US-09-594-669-2

```
  Query Match             10.9%;  Score 745.5;  DB 4;  Length 430;
  Best Local Similarity   45.7%;  Pred. No. 4e-47;
  Matches  168;  Conservative   65;  Mismatches  116;  Indels   19;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db         71 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 130

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |  |  ||  | :||:  ::| |    :||||||||||||:|| |      |: : ||:
Db        131 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 189

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |    | |    : ::::|:||| |||:||| |:  :   ||:||::| |  |:
Db        190 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 248

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                 | : :: ::||||  |    |    |     |  ||||||   | |:    |     || : :|
Db        249 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 305

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:|||||| | ::||: :|| ||:|||||||||||:  :| | |||:|:||:|||| |
Db        306 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 365

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||        |      |:           :: :
Db        366 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 416

Qy        359 GSEMNASS 366
                 || | |
Db        417 TEEMEACS 424
```

RESULT 17
US-09-595-684B-25
; Sequence 25, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: VAisberg, Eugeni

```
;  APPLICANT:  Wood, Kenneth
;  APPLICANT:  Yu, Ming
;  TITLE OF INVENTION: Human kinesins and methods of producing
;  TITLE OF INVENTION:  and purifying human kinesins
;  FILE REFERENCE: cytop036
;  CURRENT APPLICATION NUMBER: US/09/595,684B
;  CURRENT FILING DATE:  2002-06-24
;  PRIOR APPLICATION NUMBER: 09/295,612
;  PRIOR FILING DATE: 2000-04-20
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 678
;   TYPE: PRT
;   ORGANISM: Human
US-09-595-684B-25

  Query Match             10.8%;  Score 736;  DB 4;  Length 678;
  Best Local Similarity   45.6%;  Pred. No. 3.9e-46;
  Matches  166;  Conservative   62;  Mismatches  112;  Indels   24;  Gaps    8;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db        196 RICVCVRKRPLNKKETQMKDLDVITIPSKDVVMVHEPKQKVDLTRYLENQTFRFDYAFDD 255

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
Db        256 SAPNEMVYRFTAKPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 314

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
Db        315 AARDVFLMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 371

Qy        179 LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFI 238
Db        372 LQEREVKCVED-VLKLIDINSCRTSGQTSANAHSSRSHAVFQIILR--RKGKLHGKFSLI 428

Qy        239 DLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDI 298
Db        429 DLAGNERGADTSSADRQTRLEGAEINKSLLALKECIRALGRNKPHTPFRASKLTQVLRDS 488

Qy        299 FVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCI-----------NEEDD 346
Db        489 FIGENSRTCMIATISPGMASCENTLNTLRYANRVKELTVDPTAAGDVRPIMHHPPNQIDD 548

Qy        347 TNTE 350
Db        549 LETQ 552

RESULT 18
US-09-724-510-2
; Sequence 2, Application US/09724510
; Patent No. 6432659
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6432659el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1032
;  CURRENT APPLICATION NUMBER: US/09/724,510
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 09/675,227
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 319
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (154)...(183)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-724-510-2

  Query Match             10.2%;  Score 696.5;  DB 4;  Length 319;
  Best Local Similarity   45.8%;  Pred. No. 1.1e-43;
  Matches  149;  Conservative   51;  Mismatches  118;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db          1 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 60

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
Db         61 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 114

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
Db        115 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKXXXXXXXXXXXXXXXXXXXXX 174

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
Db        175 XXXXXXXXXVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 233

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
Db        234 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 293

Qy        304 KSIMIANISPTISCCEQTLNTLRYS 328
Db        294 KTCMIANISPSHVATEHTLNTLRYA 318

RESULT 19
US-09-723-216-2
; Sequence 2, Application US/09723216
; Patent No. 6436686
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6436686el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1032
;  CURRENT APPLICATION NUMBER: US/09/723,216
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 09/675,227
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 319
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (154)...(183)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-723-216-2

  Query Match             10.2%;  Score 696.5;  DB 4;  Length 319;
  Best Local Similarity   45.8%;  Pred. No. 1.1e-43;
  Matches  149;  Conservative   51;  Mismatches  118;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db          1 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 60

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
Db         61 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 114

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
Db        115 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKXXXXXXXXXXXXXXXXXXXXX 174
```

```
Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
                        :: |     |   |      | :||||||:: | :||  | |   |:|:|||||||
Db        175 XXXXXXXXXVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 233

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              || ||     ::||: :|| ||:|||||||||||:|  :    | ||| |:||:||:| |:| :
Db        234 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 293

Qy        304 KSIMIANISPTISCCEQTLNTLRYS 328
              |: |||||||:     | |||||||:
Db        294 KTCMIANISPSHVATEHTLNTLRYA 318
```

RESULT 20
US-09-675-227-2
; Sequence 2, Application US/09675227
; Patent No. 6461855
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6461855el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1032
; CURRENT APPLICATION NUMBER: US/09/675,227
; CURRENT FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (154)...(183)
; OTHER INFORMATION: Xaa = any amino acid
US-09-675-227-2

```
  Query Match             10.2%;  Score 696.5;  DB 4;  Length 319;
  Best Local Similarity   45.8%;  Pred. No. 1.1e-43;
  Matches  149;  Conservative   51;  Mismatches  118;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| ||||||      | :: : :||||::  || : |  ||:|:|| :| |   |:|| :
Db          1 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 60

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                  |  ||   |  |||    :: ||   :|||||||:||||||:|:       : ||:  |
Db         61 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 114

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              | |||   | :          ::|||||||||:|||||  :||
Db        115 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKXXXXXXXXXXXXXXXXXXXXX 174

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
                        :: |     |   |      | :||||||:: | :||  ||   |:|:|||||||
Db        175 XXXXXXXXXVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 233

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              || ||     ::||: :|| ||:|||||||||||:|  :    | ||| |:||:||:| |:| :
Db        234 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 293

Qy        304 KSIMIANISPTISCCEQTLNTLRYS 328
              |: |||||||:     | |||||||:
Db        294 KTCMIANISPSHVATEHTLNTLRYA 318
```

RESULT 21
US-09-722-129-4
; Sequence 4, Application US/09722129
; Patent No. 6582958
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6582958el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1054
; CURRENT APPLICATION NUMBER: US/09/722,129
; CURRENT FILING DATE: 2000-11-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Human
US-09-722-129-4

```
  Query Match             8.6%;  Score 589.5;  DB 4;  Length 303;
  Best Local Similarity   43.6%;  Pred. No. 8.4e-36;
  Matches  132;  Conservative   56;  Mismatches  108;  Indels    7;  Gaps    5;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |     || |||||  ::   :  :  | :  |||:|:|::    |   |  |||
Db          4 RICVCVRKRPLNQRETTLKDLDIITVPSDNVVMVHESKQKVDLTRYLQNQTFCFDHAFDD 63

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
                  |  ||: | :||:  ::  | : :|||||||||||||| |       |  : ||:
Db         64 KASNELVYQFTAQPLVESIFRKG-MATCFAYGQTGSGKTYTMGGDFSGTAQDCSKGIYAL 122

Qy        123 AAGDIFTFL--NIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
                | |:|  |   : |:| : | :: :|:||| ||:||||  :| :   ||:| ::: |  |:
Db        123 VAQDVFLLLRNSTYEKLDLK-VYGTFFEIYGGKVYDLLNWKKKLQVLEDGNQQIQVVGLQ 181

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
                      |    ||:: :   |    |            | ||||||:  | ||   :      || : :||
Db        182 EKEVCCVEEVLNLVEIGNSCRTSRQTPVNAHSSRSHAVFQIILK--SGRIMHGKFSLVDL 239

Qy        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
                  ||:||||||    ::: | :|| ||:|||||||||  |:   :| | ||| |:|| |||| |:
Db        240 AGNERGADTTKASRKRQLEGAEINKSLLALKECILALGQNKPHTPFRASKLTLVLRDSFI 299

Qy        301 GKS 303
              |::
Db        300 GQN 302
```

RESULT 22
US-09-722-129-2
; Sequence 2, Application US/09722129
; Patent No. 6582958
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6582958el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1054
; CURRENT APPLICATION NUMBER: US/09/722,129
; CURRENT FILING DATE: 2000-11-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Human
US-09-722-129-2

```
  Query Match             8.6%;  Score 589.5;  DB 4;  Length 492;
  Best Local Similarity   43.6%;  Pred. No. 1.7e-35;
  Matches  132;  Conservative   56;  Mismatches  108;  Indels    7;  Gaps    5;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |     || |||||  ::   :  :  | :  |||:|:|::    |   |  |||
Db        193 RICVCVRKRPLNQRETTLKDLDIITVPSDNVVMVHESKQKVDLTRYLQNQTFCFDHAFDD 252

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
                  |  ||: | :||:  ::  | : :|||||||||||||| |       |  : ||:
```

```
Db        253 KASNELVYQFTAQPLVESIFRKG-MATCFAYGQTGSGKTYTMGGDFSGTAQDCSKGIYAL 311

Qy        123 AAGDIFTFL--NIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180

Db        312 VAQDVFLLLRNSTYEKLDLK-VYGTFFEIYGGKVYDLLNWKKKLQVLEDGNQQIQVVGLQ 370

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240

Db        371 EKEVCCVEEVLNLVEIGNSCRTSRQTPVNAHSSRSHAVFQIILK--SGRIMHGKFSLVDL 428

Qy        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300

Db        429 AGNERGADTTKASRKRQLEGAEINKSLLALKECILALGQNKPHTPFRASKLTLVLRDSFI 488

Qy        301 GKS 303

Db        489 GQN 491
```

RESULT 23
US-09-914-259-28
; Sequence 28, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1111
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-914-259-28

```
  Query Match             8.2%;  Score 561.5;  DB 4;  Length 1111;
  Best Local Similarity   22.9%;  Pred. No. 6.1e-33;
  Matches  255;  Conservative  204;  Mismatches  393;  Indels  263;  Gaps   48;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITV---KNNCTLYIDEPRYKVDMTKYIERHEFIVD 58

Db         50 DSNIHVYVRCRSRNKREIEEKSSVVISTLGPQGKEIILSNGSHQSYSSSK--KTYQF--D 105

Qy         59 KVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLG---------S 108

Db        106 QVFGAESDQETVFNATAKNYIKEML-HGYNCTIFAYGQTGTGKTYTMSGDINILGDVQST 164

Qy        109 QPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQ--------- 159

Db        165 DNLLLGEHAGIIPRVLVDLFKELSSLNKEYS--VKISFLELYNENLKDLLSDSEDDDPAV 222

Qy        160 ---KRKM-VAALENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRS 215

Db        223 NDPKRQIRIFDNNNNNSSIMVKGMQEIFINSAHEGLNLLMQGSLKRKVAATKCNDLSSRS 282

Qy        216 HAILNI-------DLKD--INKN-TSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINR 265

Db        283 HTVFTITTNIVEQDSKDHGQNKNFVKIGKLNLVDLAGSENINRSGAENKRAQEAGL-INK 341

Qy        266 SLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTL 325

Db        342 SLLTLGRVINALVDHSNHIPYRESKLTRLLQDSLGGMTKTCIIATISPAKISMEETASTL 401

Qy        326 RYSSRVKNFKN---------KSTCINE------------EDDTNTERISILDSKGSEMNA 364

Db        402 EYATRAKSIKNTPQVNQSLSKDTCLKDYIQEIEKLRNDLKNSRNKQGIFITQ---DQLDL 458

Qy        365 SSIENVVIKSNHLLSNNNNNKINRGKIN--DKIERNNILKN---------KSFDKPREGF 413

Db        459 YESNSILIDEQNLKIHNLREQIKKFKENYLNQLDINNLLQSEKEKLIAIIQNFNVDFSNF 518

Qy        414 TSTFGK--YSSLNDIDKIKKNKKKGLINYKSTLYNDN-------------TINKKHNNNN 458

Db        519 YSEIQKIHHTNLELMNEVIQQRDFSLENSQKQ-YNTNQNMQLKISQQVLQTLNTLQGSLN 577

Qy        459 NNNNDNND---------NNNDNNNNNNNDSS-SMVNNMINHMINNNINNNI-NVNNNNNN 507

Db        578 NYNSKCSEVIKGVTEELTRNVNTHKAKHDSTLKSLLNITTNLLMNQMNELVRSISTSLEI 637

Qy        508 NNNNNNSHN----NHLPQPNYAF--------------TDTSDFSSLDDM--NCHLNNNDK 547

Db        638 FQSDSTSHYRKDLNEIYQSHQQFLKNLQNDIKSCLDSIGSSILTSINEISQNCTTNLNSM 697

Qy        548 SIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNL-----------------------HLAR- 583

Db        698 NVLIENQQSGSSKLIKEQ---DLEIKKLKNDLINERRISNQFNQQLAEMKRYFQDHVSRT 754

Qy        584 ----HSVGSKLTMFSYDPQKNKDNTFFKSN----------INKMEDNTPKDILYESRN-- 627

Db        755 RSEFHDELNKCIDNLKDKQSKLDQDIWQKTASIFNETDIVVNKIHSDSIASLAHNAENTL 814

Qy        628 --VSNMN---GNVLLGLNKNTHHDISTK----DENHDNKINNGVINIINNSNVNSINNS 678

Db        815 KTVSQNNESFTNDLISLSRGMNMDISSKLRSLPINEFLNKISQTICETCGDDNTIA-SNP 873

Qy        679 NMNSINN-SNMNSNSIYKSNYNSNQSISDVQIRYVNEMDT-SNKNNDNIFFDAISCDNNM 736

Db        874 VLTSIKKFQNIICSDIALTNEKIMSLIDEIQ----SQIETISNENNINL----------- 918

Qy        737 YPNITNNNNNNNNNN----NNNNNNIDVENYNNRDGTNNSMKLYA--------YNSHNLF 784

Db        919 ---IAINENFNSLCNFILTDYDENIMQISKTQDEVLSEHCEKLQSLKILGMDIFTAHSIE 975

Qy        785 QP---DNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKD-KNIH 840

Db        976 KPLHEHTRPEASVIKALPL---------LDYPKQFQIYR--------DAENKSKDDTSNSR 1019

Qy        841 GCDNNIIQNRN----DFEKK----------------KKTNFYNNNNIVIVNNNMG----- 875

Db       1020 TCIPNLSTNENFPLSQFSPKTPVPVPDQPLPKVLIPKSINSAKSNRSKTLPNTEGTGRES 1079

Qy        876 NNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNE 910

Db       1080 QNNLKR-------RFTTEPILKGEETENNDILQNK 1107
```

RESULT 24
US-09-541-782-6
; Sequence 6, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-09-541-782-6

```
  Query Match             7.9%;  Score 538.5;  DB 3;  Length 1073;
  Best Local Similarity   20.7%;  Pred. No. 2.9e-31;
  Matches  235;  Conservative  223;  Mismatches  410;  Indels  265;  Gaps   43;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYID-------EPRYKVDMTKYIERHEF 55
```

```
Db      59 TNINVVVRVRGRTDQE---------VRDNSSLAVSTSGAMGAELAIQSDPSSMLVTKTY 108

Qy      56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD 115
Db     109 AFDKVFGPEADQLMLFENSVAPMLEQVL-NGYNCTIFAYGQTGTGKTYTMSGD----LSD 163

Qy     116 TPGIFQYAAG----DIFTFLNIYDKDNTK-GIFISFYEIYCGKLYDLL---QKRKMVAAL 167
Db     164 SDGILSEGAGLIPRALYQLFSSLDNSNQEYAVKCSYYELYNEEIRDLLVSEELRKPARVF 223

Qy     168 EN---------GKKEVVVKD----LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESS 213
Db     224 EDTSRRGNVVITGIEESYIKNAGDGLRLLR------------EGSHRRQVAATKCNDLSS 271

Qy     214 RSHAILNIDLK-------------DINKNTS----LGKIAFIDLAGSERGADTVSQNKQ 255
Db     272 RSHSIFTITLHRKVSSGMTDETNSLTINNNSDDLLRASKLHMVDLAGSENIGRSGAENKR 331

Qy     256 TQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTI 315
Db     332 ARETGM-INQSLLTLGRVINALVEKAHHIPYRESKLTRLLQDSLGGKTKTSMIVTVSSTN 390

Qy     316 SCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISI------LDSKGSEMNASSIEN 369
Db     391 TNLEETISTLEYAARAKSIRNKP----QNNQLVFRKVLIKDLVLDIERLKNDLNATRKKN 446

Qy     370 VV------------IKSNHLLSNNNNNKINRGKINDKIERNNI--LKNKSFDKPREGFT 414
Db     447 GVYLAESTYKELMDRVQNKDLLCQEQARKLEVLDLNVKSSREQLQYVSKSNQEHKKEVEA 506

Qy     415 STFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNN----------- 462
Db     507 LQLQLVNSSTELESVKSENEK----LKNELVLEIEKRKKYETNEAKITTVATDLSQYYRE 562

Qy     463 ---------DNNDNNNDNNNNNNNDSSSMVNNMINHM--INNNINNNIN-----VNNNNN 506
Db     563 SKEYIASLYEKLDRTERNNKENENNFWNLKFNLLTMLRSFHGSFTDETNGYFTLLNDFNA 622

Qy     507 NNNNNNNSHNNHLPQPNYAFTD-TSDFSSLDD------MNCHLNNNDKSIFLHKKNLRDN 559
Db     623 SMEELLNTHSNQL---LISMTKITEHFQSLDEALQSARSSCAVPNS--SLDLIVSELKD- 676

Qy     560 IKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPK 619
Db     677 ----SKNSLLDALEHSLQDISMSSQKLGNGISSELIELQKDMKESY------------R 719

Qy     620 DILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSN 679
Db     720 QLVQELRSLYNL---------QHTHEE--------SQKELMYGVRN-----DIDALVKTC 757

Qy     680 MNSINNSNM---NSNSIYKSNYNSNQ-----SISDVQIRYVNEMDTSNKNNDNIFFDAIS 731
Db     758 TTSLNDADIILSDYISDQKSKFESKQQDLIANIGKIVSNFLQE------QNESLYTKA-- 809

Qy     732 CDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKN 791
Db     810 -------DILHSHLNDTNSNIRKANEI----------MNNRSEEFLRNA---------- 841

Qy     792 TSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCD------NN 845
Db     842 ASQAEIVGANKERIQKTVENGSQL----LDSKSKAIHSNSRSMYDHCLALAESQKQGVNL 897

Qy     846 IIQNRNDFEKKKKTNFYNNNN------IVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNN 899
Db     898 EVQTLDRLLQKVKEHSEDNTKEKHQQLLDLLESLVGNNDN------------LIDSIKTP 945

Qy     900 EMKNNEMKDNEMKDNHIKSNNNNS--SSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIR 957
Db     946 HTELQKITDHVLKGTTSLANHTNELLGLGDESLCNLETTIEDTSLVK-----LETTGDTP 1000

Qy     958 RKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNK 1010
Db    1001 SKRELPATPSWTRDSSLIKETTNLN--LDSDKKFVRETYTSSNQTNEPDVYDK 1051
```

```
RESULT 25
US-09-723-820-6
; Sequence 6, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
;  APPLICANT: Nislow, Corey
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  TITLE OF INVENTION: Antifungal Assay
;  FILE REFERENCE: 1015
;  CURRENT APPLICATION NUMBER: US/09/723,820
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: 09/541,782
;  PRIOR FILING DATE: 2000-04-03
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 1073
;   TYPE: PRT
;   ORGANISM: S.pombe
US-09-723-820-6

  Query Match             7.9%;  Score 538.5;  DB 4;  Length 1073;
  Best Local Similarity   20.7%;  Pred. No. 2.9e-31;
  Matches  235;  Conservative  223;  Mismatches  410;  Indels  265;  Gaps   43;

Qy       3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYID-------EPRYKVDMTKYIERHEF 55
Db      59 TNINVVVRVRGRTDQE---------VRDNSSLAVSTSGAMGAELAIQSDPSSMLVTKTY 108

Qy      56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD 115
Db     109 AFDKVFGPEADQLMLFENSVAPMLEQVL-NGYNCTIFAYGQTGTGKTYTMSGD----LSD 163

Qy     116 TPGIFQYAAG----DIFTFLNIYDKDNTK-GIFISFYEIYCGKLYDLL---QKRKMVAAL 167
Db     164 SDGILSEGAGLIPRALYQLFSSLDNSNQEYAVKCSYYELYNEEIRDLLVSEELRKPARVF 223

Qy     168 EN---------GKKEVVVKD----LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESS 213
Db     224 EDTSRRGNVVITGIEESYIKNAGDGLRLLR------------EGSHRRQVAATKCNDLSS 271

Qy     214 RSHAILNIDLK-------------DINKNTS----LGKIAFIDLAGSERGADTVSQNKQ 255
Db     272 RSHSIFTITLHRKVSSGMTDETNSLTINNNSDDLLRASKLHMVDLAGSENIGRSGAENKR 331

Qy     256 TQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTI 315
Db     332 ARETGM-INQSLLTLGRVINALVEKAHHIPYRESKLTRLLQDSLGGKTKTSMIVTVSSTN 390

Qy     316 SCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISI------LDSKGSEMNASSIEN 369
Db     391 TNLEETISTLEYAARAKSIRNKP----QNNQLVFRKVLIKDLVLDIERLKNDLNATRKKN 446

Qy     370 VV------------IKSNHLLSNNNNNKINRGKINDKIERNNI--LKNKSFDKPREGFT 414
Db     447 GVYLAESTYKELMDRVQNKDLLCQEQARKLEVLDLNVKSSREQLQYVSKSNQEHKKEVEA 506

Qy     415 STFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNN----------- 462
Db     507 LQLQLVNSSTELESVKSENEK----LKNELVLEIEKRKKYETNEAKITTVATDLSQYYRE 562

Qy     463 ---------DNNDNNNDNNNNNNNDSSSMVNNMINHM--INNNINNNIN-----VNNNNN 506
Db     563 SKEYIASLYEKLDRTERNNKENENNFWNLKFNLLTMLRSFHGSFTDETNGYFTLLNDFNA 622

Qy     507 NNNNNNNSHNNHLPQPNYAFTD-TSDFSSLDD------MNCHLNNNDKSIFLHKKNLRDN 559
Db     623 SMEELLNTHSNQL---LISMTKITEHFQSLDEALQSARSSCAVPNS--SLDLIVSELKD- 676

Qy     560 IKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPK 619
```

```
Db        677 ----SKNSLLDALEHSLQDISMSSQKLGNGISSELIELQKDMKESY------------R 719

Qy        620 DILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSN 679
Db        720 QLVQELRSLYNL---------QHTHEE--------SQKELMYGVRN-----DIDALVKTC 757

Qy        680 MNSINNSNM---NSNSIYKSNYNSNQ-----SISDVQIRYVNEMDTSNKNNDNIFFDAIS 731
Db        758 TTSLNDADIILSDYISDQKSKFESKQQDLIANIGKIVSNFLQE------QNESLYTKA-- 809

Qy        732 CDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKN 791
Db        810 -------DILHSHLNDTNSNIRKANEI----------MNNRSEEFLRNA----------- 841

Qy        792 TSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCD------NN 845
Db        842 ASQAEIVGANKERIQKTVENGSQL----LDSKSKAIHSNSRSMYDHCLALAESQKQGVNL 897

Qy        846 IIQNRNDFEKKKKTNFYNNNN------IVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNN 899
Db        898 EVQTLDRLLQKVKEHSEDNTKEKHQQLLDLLESLVGNNDN------------LIDSIKTP 945

Qy        900 EMKNNEMKDNEMKDNHIKSNNNNS--SSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIR 957
Db        946 HTELQKITDHVLKGTTSLANHTNELLGLGDESLCNLETTIEDTSLVK-----LETTGDTP 1000

Qy        958 RKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNK 1010
Db       1001 SKRELPATPSWTRDSSLIKETTNLN--LDSDKKFVRETYTSSNQTNEPDVYDK 1051
```

RESULT 26
US-09-417-485D-6
; Sequence 6, Application US/09417485D
; Patent No. 6541202
; GENERAL INFORMATION:
; APPLICANT: Long, David M.
; APPLICANT: Metz, Anneke M.
; APPLICANT: Love, Ruschelle A.
; TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
; FILE REFERENCE: 47714-5009-US
; CURRENT APPLICATION NUMBER: US/09/417,485D
; CURRENT FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 2184
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (330)..(335)
;   OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
;   OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
US-09-417-485D-6

```
  Query Match             7.9%;  Score 538.5;  DB 4;  Length 2184;
  Best Local Similarity   21.4%;  Pred. No. 8e-31;
  Matches  330;  Conservative  234;  Mismatches  489;  Indels  487;  Gaps   73;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLY----IDEPRYKVDMTKYIERH--EFI- 56
Db        429 QIKCSTKKNDISHIITSRKENHLFHVQKLENNYKHPNINKQLRKTKILKYVYNYFKEFIN 488

Qy         57 ----------------------VDKVF---------DDTVDNFTVYENTIKPLIIDLYE 84
Db        489 NVINTKFGKIYRKFFPRKHILNKIHKIFKIIRLQIIKKYRIINIRMNRKFIKQKVYDTFF 548

Qy         85 NGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFI 144
Db        549 KNYDFLSFSF------KTYKIINFMVYITKKCIPKLLGSKHNF---KIFLKNVKKFLLF 599

Qy        145 SFYEIYC-------GKLYDLLQKRKMVAALENGKKEVVVKDL-------KILRVLTKEEL 190
Db        600 NYKESFSLNQVMKNIKVKNIFQKK---ISKYNIKNRILLKNIFDNNYENKILH-RNNKEI 655

Qy        191 ILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADT- 249
Db        656 ITNINDNIKI----YNKKNDNLNNSFKIKTTLFNKLRRK-YFNKIKKINIAIQKRHLMNR 710

Qy        250 -------------------VSQNKQTQTDGANINRSLLALKECIRAMDSDKN----HIPF 286
Db        711 LIYFLFNYFIMPLIRRFFFLTKSEQT------LHKTIFFDRKIWNHFTKISNFCLYHQIF 764

Qy        287 RDSELTK-------VLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKST 339
Db        765 RNKKLKKRNEPKMDYVQNMFNVKKKG-------------EKIKTNKYIF-IKKMKKKST 809

Qy        340 --CINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRG-------- 389
Db        810 NKCINNK---FSKKCIPKKKKKNLYNITRHNNIFIKKDMEKKSKTNNLINKSIDNLYKLK 866

Qy        390 KINDKI------------------------------ERNNILKNKSF----------DK 408
Db        867 EINKKSVRPYIKKFYYKIKKKYFALKKMYIHMRMAKEEKSNIKLERAFKHFFIFAQEKEH 926

Qy        409 PREGFTSTFGKYSSLND-------IDKIK-----------KNKKK---GLINYKSTLYND 447
Db        927 ILKYFSSHFFQNRKINYGKRFNKLIHRIKNIIIKQNSGIVKNKDKTFLHLIKNKSNKNNN 986

Qy        448 NTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNN 507
Db        987 NKKKNKNNYNNNNINNNNNNNNNNNINNNNNNKCKLSN----------SKRYNIRNNNNN 1036

Qy        508 NNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSS 567
Db       1037 KKAKNNEKNN----------------IDDSN--LEKKKKKIYIY--------KIKNIIE 1069

Qy        568 CDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPK-------D 620
Db       1070 KRNFMLK----LNSINHFISKKLRI-NWIPKKKG----LRPLINLSTLNVPEIVKQRIFE 1120

Qy        621 ILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNM 680
Db       1121 ILKSKKSSEFYFHNILNNLER------EKKDKNIKKRKKYNK--KNFNPVSLNNICNFSL 1172

Qy        681 NSINNSNMNSNSIYKSNYNSNQSISDVQI-----------------RYVNEMDTSNKNND 723
Db       1173 KCLGNMRHNNNSLFK---NTLTKTGEIELKLKKWLHYLKNWFYKKKRMKKYIKNKLKNNK 1229

Qy        724 NIF-FDAISCDNNMYPNITNNNNNNN------NNNNNNNNNIDVENYNNRDGTNNSMKLY 776
Db       1230 KIYAYICIGDFSNCYEHI-----NHNYLFKILKNFFDNINNFEFIYLFKR-----SFRLY 1279

Qy        777 AYNSHNLF---QPDNNKN--TSNIQNIN--TNKNNQDGNVNYSMN--FCHYNLNDKNYLI 827
Db       1280 NKNLNNSFLSYYPVNVKSFGLHYIRNLRELIIKSHLNDNHHFLLNQMFKTKSKSDLYIFA 1339

Qy        828 DLNNKEQKDKN---------IHGCDNNIIQNRNDFEKKKKTNFY------NNNNIVIVN- 871
Db       1340 DSYKSLQVDKRDIFMTIITVIRYYYLNIYFSIKEFKLNRKNIFYFQIFQENQMKGVYLSV 1399

Qy        872 -------------NNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIK 917
Db       1400 RDKKRVENIKKWYLNSMKKINHDEILESLKNSSININN------KNFMICTNHEQDTEEK 1453

Qy        918 SNNNNSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYT 977
Db       1454 GNTQNKEKHDIYIGPIYNNSFDSTT----TTHSSNNY---KGNNIHVSGD-YKNDGLLHK 1505

Qy        978 INSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNN 1037
Db       1506 GNN---------------------SMNE-CYVKDIK------------CNNNNNNNNNNN 1531

Qy       1038 KNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDT 1097
```

```
Db       1532 NNNINNSY---NKLNCVTNNSKNDI-----------IKYHKTIDTDNSKNHTYFK----- 1572

Qy       1098 IINNCLNSLD---ISSMYDDTKEI-LNNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDK 1153

Db       1573 --NKFLNFLDKKIISNIYGLPQGFSLSNILCSLYYAYLDK------NEEFQNL------- 1617

Qy       1154 TAQSIYEKRKVLLTKLLLLFKKNVDTQINNE------TSDLRKDLVMCHICNNNP----- 1202

Db       1618 ----LYSEK----------------QINNKYFLANGT---------CNYFNLNSLILRF 1647

Qy       1203 DDQFHFYAYSRLEKDII-NLIMLRQIWCESEN--------LRLLYQ--FLVVEYQNKSAN 1251

Db       1648 IDDFLFITLNKKNIKIFKNLLLKKKIWGSNINSSKTKIFKIPLIYKNDLLIYNFQNKYQK 1707

Qy       1252 SVLLNVSSNNGDIILLNKKLVQDNIKNSMDHN---NIHKK 1288

Db       1708 K----------KYKIKNKKKIQ-SVRNKRIHNQLVNANKK 1736
```

RESULT 27
US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
;   APPLICANT: Stewart, Thomas S
;   APPLICANT: Flores, Maria V
;   APPLICANT: O'Sullivan, William J
;   TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
;   TITLE OF INVENTION: phosphate synthetase II
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Nixon & Vanderhye PC
;   STREET: 1100 No. 5849573th Glebe Road, 8th Floor
;   CITY: Arlington
;   STATE: Virginia
;   COUNTRY: USA
;   ZIP: 22201-4714
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.24
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/446,855A
;   FILING DATE: 06-Jul-1995
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Mitchard, Leonard C
;   REGISTRATION NUMBER: 29.009
;   REFERENCE/DOCKET NUMBER: 47-80
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 703-816-4000
;   TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2391 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-446-855A-2

Query Match 7.8%; Score 533.5; DB 2; Length 2391;
Best Local Similarity 20.4%; Pred. No. 2.1e-30;
Matches 272; Conservative 225; Mismatches 404; Indels 431; Gaps 57;

```
Qy        189 ELILKMID-----GVLLRKIGVNSQ---------NDESSRSHAILN-------------- 220

Db        948 EIALKVITHLNIIGECNIQFGINPQTGEYCIIEVNARLSRSSALASKATGYPLAYISAKI 1007

Qy        221 ------IDLKD-INKNTS----------LGKIAFIDLAGSERGADTVSQNKQTQTDGANI 263

Db       1008 ALGYDLISLKNSITKKTTACFEPSLDYITTKIPRWDLNKFEFASNTMNSSMKSVGEVMSI 1067

Qy        264 NRSL-LALKECIRAMDSDKNHIPFR-----DSELTKVLRDIFVGKSKSIMIANISP-TIS 316

Db       1068 GRTFEESIQKSLRCIDD--NYLGFSNTYCIDWDEKKIIEE----------LKNPSPKRID 1115

Qy        317 CCEQTLNT-----------------LRYSSRVKNFKNKSTCINEEDDTNTERISILDSKG 359

Db       1116 AIHQAFHLNMPMDKIHELTHIDYWFLHKFYNIYNLQNKLKTL------KLEQLSFNDLKY 1169

Qy        360 SEMNASSIENVVIKSNHLLS------NNNNNKINRGKINDKIERNNILKNKSFDKPREGF 413

Db       1170 FKKHGFSDKQIA----HYLSFNTSDNNNNNNNISSCRVTE----NDVMK----------Y 1211

Qy        414 TSTFGKYSSLNDIDKI-------------------------------------KKNKKKG 436

Db       1212 REKLGLFPHIKVIDTLSAEFPALTNYLYLTYQGQEHDVLPLNMKRKKICTLNNKRNANKK 1271

Qy        437 LINYKSTLYN------DNTINKKHNNNNNNNNDNNDN-----------NNDNNNNNNNDS 479

Db       1272 KVHVKNHLYNEVVDDKDTQLHKENNNNNNMSGNVENKCKLNKESYGYNNSSNCINTNNI 1331

Qy        480 SSMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMN 539

Db       1332 N--IENNICHDISINKNIKVTINNSNNSISNNENVETN-------------------LN 1369

Qy        540 CHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQK 599

Db       1370 C--------------------VSERAGSHHIYGKEEK-------SIGSDDT--NILSAQ 1399

Qy        600 NKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKI 659

Db       1400 NSNNNFSCNN-------------ENMNKANVDVNVL-------ENDTKKREDINTTTVF 1438

Qy        660 NNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSN 719

Db       1439 MEGQNSVINNKNKE--NSSLLKGDEEDIVMVNLKKENNYNS----------VINNVDCRK 1486

Qy        720 KNND--NIFFDAISCDNNMYPNITNNNN----NNNNNNNNNNNIDVENYNNRD---GTN 770

Db       1487 KDMDGKNINDECKTYKKNKYKDMGLNNNIVDELSNGTSHSTNDHLYLDNFNTSDEEIGNN 1546

Qy        771 NSMKLYAYNSHNLFQPDNNKNTSN----------------IQNINTNKN-NQDGNVNYS 812

Db       1547 KNMDMYLSKEKSI----SNKNPGNSYYVVDSVYNNEYKINKMKELIDNENLNDEYNNNVN 1602

Qy        813 MNFCHYNLNDKNYLIDLNNKEQKDKNIHGC-DNNIIQNRNDFEKKKKTNFYNNNNIVIVN 871

Db       1603 MNCSNYN----NASAFVNGKDRNDNLENDCIEKNMDHTYKHYNRLNNRRSTNERMMLMVN 1658

Qy        872 NNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNN 931

Db       1659 NEKESNHEKGHRRNGLNKKNKEKNMEKNKGKNKD-KKNYHYVNHKRNNEYNSNNIESKFN 1717

Qy        932 NIYNNINDDDTFQNDYCHND-NTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLL 990

Db       1718 NYVDDINKKEYYEDE---NDIYYFTHSSQGN---------NDDL-----SNDNYLSSEEL 1760

Qy        991 HFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDD--NDNNNVDNNNKNNVDN---NV 1045

Db       1761 NTDE--------YDDDYYYDE---------DEEDDYDDNDDDDDDDDGEDEEDNDYYND 1804

Qy       1046 DNNNVDNNVDNNDKNNVD------NNNVDNDD-DDVDFHNIKNFNNNE------------ 1086

Db       1805 DGYDSYNSLSSSRISDVSSVIYSGNENIFNEKYNDIGFKIIDNRNEKEKEKKKCFIVLGC 1864

Qy       1087 ------------------YLSYFQKNVDTIINNCLNSLDISSMYDDT-----KEILNNIL 1123

Db       1865 GCYRIGSSVEFDWSAIHCVKTIRKLNHKAILINC-NPETVSTDYDESDRLYFDEITTEVI 1923

Qy       1124 LSKYKAEKDN-VIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQIN 1182

Db       1924 KFIYNFENSNGVIIAFGGQTSNNL-----------------------VFSLYKNNV----- 1956
```

```
Qy       1183 NETSDLRKDLVMCHICNNNPD------DQFHFYAYSRLEKDI-----------------I 1219
               |   : | |::  |  |          ||  :  :::| | |                  :
Db       1957 NILGSVHKVLIVVKIGINFRTYVILKIDQPKWNKFTKLSKAIQFANEVKFPVLVRPSYVL 2016

Qy       1220 NLIMLRQIWCESENLRLLYQFLVVEYQN--------KSANSVLLNVSSNNGDIILLNKKL 1271
               :    :| :  |   |     |  :  :|   |          ::|  : ::  | || ||
Db       2017 SGAAMRVVNCFEELKNFLMKAAIVSKDNPVVISKFIENAKEIEIDCVSKNGKII---NYA 2073

Qy       1272 VQDNIKNSMDHN 1283
               : ::::|:  |:
Db       2074 ISEHVENAGVHS 2085
```

RESULT 28
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 2391
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

```
  Query Match             7.8%;  Score 533.5;  DB 3;  Length 2391;
  Best Local Similarity   20.4%;  Pred. No. 2.1e-30;
  Matches  272;  Conservative  225;  Mismatches  404;  Indels  431;  Gaps   57;

Qy        189 ELILKMID-----GVLLRKIGVNSQ---------NDESSRSHAILN-------------- 220
              |: ||:|        |     :|:|          |     |||  |: :
Db        948 EIALKVITHLNIIGECNIQFGINPQTGEYCIIEVNARLSRSSALASKATGYPLAYISAKI 1007

Qy        221 ------IDLKD-INKNTS----------LGKIAFIDLAGSERGADTVSQNKQTQTDGANI 263
                    | ||: | | |:          ||    ||    |   ::|:: : ::  :  :|
Db       1008 ALGYDLISLKNSITKKTTACFEPSLDYITTKIPRWDLNKFEFASNTMNSSMKSVGEVMSI 1067

Qy        264 NRSL-LALKECIRAMDSDKNHIPFR-----DSELTKVLRDIFVGKSKSIMIANISP-TIS 316
              |:   :::: :| :|   |:: |       | :  |:: :             : | || |
Db       1068 GRTFEESIQKSLRCIDD--NYLGFSNTYCIDWDEKKIIEE----------LKNPSPKRID 1115

Qy        317 CCEQTLNT-----------------LRYSSRVKNFKNKSTCINEEDDTNTERISILDSKG 359
                  |  :                  |        :  | :||    :        |::|  | |
Db       1116 AIHQAFHLNMPMDKIHELTHIDYWFLHKFYNIYNLQNKLKTL------KLEQLSFNDLKY 1169

Qy        360 SEMNASSIENVVIKSNHLLS------NNNNNKINRGKINDKIERNNILKNKSFDKPREGF 413
                 : :   | : :      | ||     |||||  |:   :: :     |:::|          :
Db       1170 FKKHGFSDKQIA----HYLSFNTSDNNNNNNNISSCRVTE----NDVMK----------Y 1211

Qy        414 TSTFGKYSSLNDIDKI--------------------------------------KKNKKKG 436
                   | :   :  || :                                      |:|  |
Db       1212 REKLGLFPHIKVIDTLSAEFPALTNYLYLTYQGQEHDVLPLNMKRKKICTLNNKRNANKK 1271

Qy        437 LINYKSTLYN------DNTINKKHNNNNNNNNDNNDN------------NNDNNNNNNDS 479
               :: |: |||       |   ::|::||||| |: |  :|              || :|  | |:
Db       1272 KVHVKNHLYNEVVDDKDTQLHKENNNNNNMNSGNVENKCKLNKESYGYNNSSNCINTNNI 1331

Qy        480 SSMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMN 539
              :  :  | |  | |  : |  |   :  :||:||: :|| |     |              :|
Db       1332 N--IENNICHDISINKNIKVTINNSNNSISNNENVETN--------------------LN 1369

Qy        540 CHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQK 599
              |                    :  |:   :|  |::         |:||  |   :    :
Db       1370 C---------------------VSERAGSHHIYGKEEK-------SIGSDDT--NILSAQ 1399

Qy        600 NKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKI 659
              | :|  |   :|                |:  | :|:: |||            :|    ::: :
Db       1400 NSNNNFSCNN--------------ENMNKANVDVNVL-------ENDTKKREDINTTTVF 1438

Qy        660 NNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSN 719
                  |  ::|||  |     |:| :           :   |     ::||||            :| :|
Db       1439 MEGQNSVINNKNKE--NSSLLKGDEEDIVMVNLKKENNYNS----------VINNVDCRK 1486

Qy        720 KNND--NIFFDAISCDNNMYPNITNNNN----NNNNNNNNNNNIDVENYNNRD---GTN 770
              |: |  ||   :   :    | |  ::  |||       :|    :::  |:: ::|:|  |   | |
Db       1487 KDMDGKNINDECKTYKKNKYKDMGLNNNIVDELSNGTSHSTNDHLYLDNFNTSDEEIGNN 1546

Qy        771 NSMKLYAYNSHNLFQPDNNKNTSN------------------IQNINTNKN-NQDGNVNYS 812
               :| :|        ::     :|||  |                   :: :   |:|  | :  |  | :
Db       1547 KNMDMYLSKEKSI----SNKNPGNSYYVVDSVYNNEYKINKMKELIDNENLNDEYNNNVN 1602

Qy        813 MNFCHYNLNDKNYLIDLNNKEQKDKNIHGC-DNNIIQNRNDFEKKKKTNFYNNNNIVIVN 871
              || :||     |       :| |::  |     :  |  : |:          : :         |    :::||
Db       1603 MNCSNYN----NASAFVNGKDRNDNLENDCIEKNMDHTYKHYNRLNNRRSTNERMMLMVN 1658

Qy        872 NNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNNSSSSSSNN 931
               |    :|:     :              ||:  |:  || :  |  |     ||  ::|   ||::  |  |
Db       1659 NEKESNHEKGHRRNGLNKKNKEKNMEKNKGKNKD-KKNYHYVNHKRNNEYNSNNIESKFN 1717

Qy        932 NIYNNINDDDTFQNDYCHND-NTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLL 990
               |   ::||  : ::::     ||    ||      |            |||:     | ::|:|:  |
Db       1718 NYVDDINKKEYYEDE---NDIYYFTHSSQGN----------NDDL-----SNDNYLSSEEL 1760

Qy        991 HFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDD--NDNNNVDNNNKNNVDN---NV 1045
               :  |        ::| |  |          |::| |||  :|::: |:: ::  ||   |
Db       1761 NTDE---------YDDDYYYDE---------DEEDDYDDDNDDDDDDDDGEDEEDNDYYND 1804

Qy       1046 DNNNVDNNVDNNDKNNVD------NNNVDNDD-DDVDFHNIKNFNNNE------------ 1086
               |  :  |:: ::  ::|        | |: |:  :|: |  | | | |
Db       1805 DGYDSYNSLSSSRISDVSSVIYSGNENIFNEKYNDIGFKIIDNRNEKEKEKKKCFIVLGC 1864

Qy       1087 ------------------YLSYFQKNVDTIINNCLNSLDISSMYDDT-----KEILNNIL 1123
                                  :  :  |    |: ||   |   :|: ||::       |   ::
Db       1865 GCYRIGSSVEFDWSAIHCVKTIRKLNHKAILINC-NPETVSTDYDESDRLYFDEITTEVI 1923

Qy       1124 LSKYKAEKDN-VIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQIN 1182
                    |   |   |  ||   :   :       |:                         :  |:| ||
Db       1924 KFIYNFENSNGVIIAFGGQTSNNL----------------------VFSLYKNNV----- 1956

Qy       1183 NETSDLRKDLVMCHICNNNPD------DQFHFYAYSRLEKDI-----------------I 1219
                 :  | |:::  |  |          ||  :  :::| | |                  :
Db       1957 NILGSVHKVLIVVKIGINFRTYVILKIDQPKWNKFTKLSKAIQFANEVKFPVLVRPSYVL 2016

Qy       1220 NLIMLRQIWCESENLRLLYQFLVVEYQN--------KSANSVLLNVSSNNGDIILLNKKL 1271
               :    :| :  |   |     |  :  :|   |          ::|  : ::  | || ||
Db       2017 SGAAMRVVNCFEELKNFLMKAAIVSKDNPVVISKFIENAKEIEIDCVSKNGKII---NYA 2073

Qy       1272 VQDNIKNSMDHN 1283
               : ::::|:  |:
Db       2074 ISEHVENAGVHS 2085
```

RESULT 29
US-09-914-259-21
; Sequence 21, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul

```
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
;   LENGTH: 967
;   TYPE: PRT
;   ORGANISM: Loligo pealeii
US-09-914-259-21

  Query Match              7.3%;  Score 500.5;  DB 4;  Length 967;
  Best Local Similarity   22.1%;  Pred. No. 1.6e-28;
  Matches  217;  Conservative  189;  Mismatches  403;  Indels  175;  Gaps   35;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 64
Db          9 IKVICRVRPLNEAEERAGSKFIL-------------KFPTDDSISIAGKVFVFDKVLKPN 55

Qy         65 VDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTP---GIFQ 121
Db         56 VSQEYVYNVGAKPIVADVL-SGCNGTIFAYGQTSSGKTHTM-----EGVLDKPSMHGIIP 109

Qy        122 YAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKI 181
Db        110 RIVQDIFNYIYGMDENLEFHIKISYYEIYLDKIRDLLDVTKTNLAVHEDKNRVPFVKGAT 169

Qy        182 LRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTS---LGKIAF 237
Db        170 ERFVSSPEEVMEVIDEGKNNRHVAVTNMNEHSSRSHSVFLINVKQENVETQKKLSGKLYL 229

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM-DSDKNHIPFRDSELTKVLR 296
Db        230 VDLAGSEKVSKTGAEG-AVLDEAKNINKSLSALGNVISALADGNKSHVPYRDSKLTRILQ 288

Qy        297 DIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEE---------DDT 347
Db        289 ESLGGNARTTMVICCSPASYNESETKSTLLFGQRAKTIKN-VVSVNEELTADEWKRRYEK 347

Qy        348 NTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKI---ERNNILKNK 404
Db        348 EKERVTKLKATMAKLEA--------------------ELQRWRTGQAVSVEEQVDLKEDV 387

Qy        405 SFDKPREGFTS-TFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKK---------- 453
Db        388 PAESPATSTTSLAGGLIASMNEGDRTQLEEER-LKLYQQLDDKDDEINNQSQLIEKLKEQ 446

Qy        454 ---------HNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNN 504
Db        447 MMEQEDLIAQSRRDYENLQQDMSRIQADNESAKDEVKEVLQALEELAMNYDQKSQEVEDK 506

Qy        505 NNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNI--KL 562
Db        507 NKENENLSEELNQKL-------------STLNSLQNELDQLKDSSMHHRKRVTDMMINLL 553

Qy        563 KNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDIL 622
Db        554 KDLGDIGTIVG---GNAAETKPTAGS-------GEKIEEEFTVARLYISKMKS------- 596

Qy        623 YESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNS 682
Db        597 -EVKTLVSRNNQL-----ENTQQDNFKKIETHEKDLSNCKLLIQQHEAKMASLQEAIKDS 650

Qy        683 INNSNMNSNSIYKSNYNSNQSISDVQIRYV----NEMDTSNKNNDNIFFDAISCDNNMYP 738
Db        651 ENKKRMLEDNVDSLNEEYAKLKAQEQMHLAALSEREKETSQASETR---EVLEKQMEMHR 707

Qy        739 NITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSM---KLYAYNSHNLFQPDNNKNTSNI 795
Db        708 E-QHQKQLQSLRDEISEKQATVDNLKD-DNQRLSLALEKLQA--DYDKLKQEEVEKAAKL 763

Qy        796 QNINTNKNNQ----------DGNVNYSMNFCHYNLNDKNYLIDLNNK--------EQKDK 837
Db        764 ADLSLQIDRREQAKQDLKGLEETVAKELQTLH-NLR-KLFVQDLQNKVKKSCSKTEEEDE 821

Qy        838 NIHGCDNNIIQNRNDFEKKKKTNFYNNN--NIVIVNNNMGNNNS------PRMKYGLCGS 889
Db        822 DTGG----------NAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRL--- 868

Qy        890 HTSIDNMKNNEMKNNEMKDNEMKD 913
Db        869 RATMERVKSLESALKDAKEGAMRD 892
```

```
RESULT 30
US-09-541-782-2
; Sequence 2, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
;  APPLICANT:  Nislow, Corey
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  TITLE OF INVENTION:  Antifungal Assay
;  FILE REFERENCE:  1015
;  CURRENT APPLICATION NUMBER: US/09/541,782
;  CURRENT FILING DATE:  2000-04-03
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1184
;   TYPE: PRT
;   ORGANISM: A. nidulans
US-09-541-782-2

  Query Match              7.3%;  Score 500;  DB 3;  Length 1184;
  Best Local Similarity   21.1%;  Pred. No. 2.4e-28;
  Matches  253;  Conservative  194;  Mismatches  415;  Indels  338;  Gaps   48;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDM-TKYIERHEFIVDKV 60
Db         79 DTSIHVVVRCRGRNEREVKENSGVVLQTEG-----VKGKTVELSMGPNAVSNKTYTFDKV 133

Qy         61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120
Db        134 FSAAADQITVYEDVVLPIVTEMLA-GYNCTIFAYGQTGTGKTYTMSGD----MTDTLGIL 188

Qy        121 QYAAGDI-----FTFLNIYDKDNTKGIFISFYEIYCGKLYDLL-------------QKR 161
Db        189 SDNAGIIPRVLYSLFAKLADTEST--VKCSFIELYNEELRDLLSAEENPKLKIYDNEQKK 246

Qy        162 KMVAALENGKKEVVVKD----LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHA 217
Db        247 GHMSTLVQGMEETYIDSATAGIKLLQ-----------QGSHKRQVAATKCNDLSSRSHT 294

Qy        218 ILNIDLKDINKNTSL-------GKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLAL 270
Db        295 VFTITV-NIKRTTESGEEYVCPGKLNLVDLAGSENIGRSGAENKRA-TEAGLINKSLLTL 352

Qy        271 KECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSR 330
Db        353 GRVINALVDKSQHIPYRESKLTRLLQDSLGGRTKTCIIATMSPARSNLEETISTLDYAFR 412

Qy        331 VKNFKNKSTCIN--------------EEDDTNTERISILDSKGSEMNASSIENVVI--KS 374
Db        413 AKNIRNKPQ-INSTMPKMTLLREFTAEIEKLKAELIATRHRNGVYMSVESYEEMKMENES 471

Qy        375 NHLLSNNNNNKIN--RGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDI----DK 428
Db        472 RRIISEEQRAKIESMESSLRHKVQELLTLTSK-FNDLKKDNDDTLAALCSTNDVLQQTDI 530

Qy        429 IKKNKK-----------------------KGLIN-YKSTLYNDNTINKKHNNNNNNNND 463
Db        531 VLQNTRAQLEEEEMLRCAHEETEHQLQDVGKGLISTLGQTVEDINSLQSK--------LD 582
```

```
Qy        464 NNDNNNDNNNNNNNDSSSMVNNMINHMINNNI-----------NNNINVN-------NN 504
Db        583 RKAELDATNAELWRASSTEVSD-VTKRIDQRVEAFQTRHAKLLETTSVKVNEFIATEISN 641
Qy        505 NNNNNNNNNSHNNHLPQP-NYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLK 563
Db        642 IERTRSDLSEYNRSLDAACNNAKAETS--SAHEDMN--------NVLEEIKDLREEVKSK 691
Qy        564 NRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILY 623
Db        692 --------VGEGLNGLSAAAARISEEVI---------GEFTQLHSQLHTSFNNLGKDL-- 732
Qy        624 ESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSI 683
Db        733 ----------------KSIFETMAT-HLSEQKNEINR------LRAELQSSNRQNIETT 768
Qy        684 NNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDA--------ISCDNN 735
Db        769 HKASAHLAQAIEEEHVAAEAEREILMSQIKALVEESRQKQFARLRAKIDGVRTEISASGD 828
Qy        736 MYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNI 795
Db        829 MLEQAT---------TQHDRQIDEWVF-------------------------KSEQFA 852
Qy        796 QNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEK 855
Db        853 KDVNASKD-----------------------EIRTKLQNDWEAFDQRNSTI-------- 880
Qy        856 KKKTNFYNNNNIVIVN---NNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMK 912
Db        881 RKATESVHKETVRIVDVQVDDMGRQME-----------ALDDFVAKARSQN----GRYR 924
Qy        913 DNHIKSNNNNSS----SSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIR--------RKN 960
Db        925 DAHIATLDTIATGVRDSYSSIEGRVENLTGRMNQFQQEATHHHATLEESIAPLSNDVRKP 984
Qy        961 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTY----PTLSTNEDIYN-----KE 1011
Db        985 LTDLSSS-FQN-------RSLEEYVATGVTPKKRKYDYISVLPSTESHEVLKSRLRTTKE 1036
Qy       1012 MEGKHIRLDDQDK--------------YDDND-----------NNNVDNNNKNNVDNNV 1045
Db       1037 MEVLPFNSDDQLSGPSSSPGGSPSKGFVYNDVEDEVGTHAPTVTNVNPSNTGLREVDANV 1096
```

RESULT 31
US-09-723-820-2
; Sequence 2, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: A. nidulans
US-09-723-820-2

```
  Query Match             7.3%;  Score 500;  DB 4;  Length 1184;
  Best Local Similarity   21.1%;  Pred. No. 2.4e-28;
  Matches  253;  Conservative  194;  Mismatches  415;  Indels  338;  Gaps   48;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDM-TKYIERHEFIVDKV 60
Db         79 DTSIHVVVRCRGRNEREVKENSGVVLQTEG-----VKGKTVELSMGPNAVSNKTYTFDKV 133
Qy         61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120
Db        134 FSAAADQITVYEDVVLPIVTEMLA-GYNCTIFAYGQTGTGKTYTMSGD----MTDTLGIL 188
Qy        121 QYAAGDI-----FTFLNIYDKDNTKGIFISFYEIYCGKLYDLL-------------QKR 161
Db        189 SDNAGIIPRVLYSLFAKLADTEST--VKCSFIELYNEELRDLLSAEENPKLKIYDNEQKK 246
Qy        162 KMVAALENGKKEVVVKD----LKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHA 217
Db        247 GHMSTLVQGMEETYIDSATAGIKLLQ-----------QGSHKRQVAATKCNDLSSRSHT 294
Qy        218 ILNIDLKDINKNTSL-------GKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLAL 270
Db        295 VFTITV-NIKRTTESGEEYVCPGKLNLVDLAGSENIGRSGAENKRA-TEAGLINKSLLTL 352
Qy        271 KECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSR 330
Db        353 GRVINALVDKSQHIPYRESKLTRLLQDSLGGRTKTCIIATMSPARSNLEETISTLDYAFR 412
Qy        331 VKNFKNKSTCIN-------------EEDDTNTERISILDSKGSEMNASSIENVVI--KS 374
Db        413 AKNIRNKPQ-INSTMPKMTLLREFTAEIEKLKAELIATRHRNGVYMSVESYEEMKMENES 471
Qy        375 NHLLSNNNNNKIN--RGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDI----DK 428
Db        472 RRIISEEQRAKIESMESSLRHKVQELLTLTSK-FNDLKKDNDDTLAALCSTNDVLQQTDI 530
Qy        429 IKKNKK-----------------------KGLIN-YKSTLYNDNTINKKHNNNNNNNND 463
Db        531 VLQNTRAQLEEEEMLRCAHEETEHQLQDVGKGLISTLGQTVEDINSLQSK--------LD 582
Qy        464 NNDNNNDNNNNNNNDSSSMVNNMINHMINNNI-----------NNNINVN-------NN 504
Db        583 RKAELDATNAELWRASSTEVSD-VTKRIDQRVEAFQTRHAKLLETTSVKVNEFIATEISN 641
Qy        505 NNNNNNNNNSHNNHLPQP-NYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLK 563
Db        642 IERTRSDLSEYNRSLDAACNNAKAETS--SAHEDMN--------NVLEEIKDLREEVKSK 691
Qy        564 NRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILY 623
Db        692 --------VGEGLNGLSAAAARISEEVI---------GEFTQLHSQLHTSFNNLGKDL-- 732
Qy        624 ESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSI 683
Db        733 ----------------KSIFETMAT-HLSEQKNEINR------LRAELQSSNRQNIETT 768
Qy        684 NNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDA--------ISCDNN 735
Db        769 HKASAHLAQAIEEEHVAAEAEREILMSQIKALVEESRQKQFARLRAKIDGVRTEISASGD 828
Qy        736 MYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNI 795
Db        829 MLEQAT---------TQHDRQIDEWVF-------------------------KSEQFA 852
Qy        796 QNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEK 855
Db        853 KDVNASKD-----------------------EIRTKLQNDWEAFDQRNSTI-------- 880
Qy        856 KKKTNFYNNNNIVIVN---NNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMK 912
Db        881 RKATESVHKETVRIVDVQVDDMGRQME-----------ALDDFVAKARSQN----GRYR 924
Qy        913 DNHIKSNNNNSS----SSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIR--------RKN 960
Db        925 DAHIATLDTIATGVRDSYSSIEGRVENLTGRMNQFQQEATHHHATLEESIAPLSNDVRKP 984
Qy        961 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTY----PTLSTNEDIYN-----KE 1011
Db        985 LTDLSSS-FQN-------RSLEEYVATGVTPKKRKYDYISVLPSTESHEVLKSRLRTTKE 1036
```

```
Qy     1012 MEGKHIRLDDQDK--------------YDDND----------NNNVDNNNKNNVDNNV 1045
Db     1037 MEVLPFNSDDQLSGPSSSPGGSPSKGFVYNDVEDEVGTHAPTVTNVNPSNTGLREVDANV 1096

RESULT 32
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Ohashi, Cara
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  VAisberg, Eugeni
;  APPLICANT:  Wood, Kenneth
;  APPLICANT:  Yu, Ming
;  TITLE OF INVENTION: Human kinesins and methods of producing
;  TITLE OF INVENTION:  and purifying human kinesins
;  FILE REFERENCE: cytop036
;  CURRENT APPLICATION NUMBER: US/09/595,684B
;  CURRENT FILING DATE:  2002-06-24
;  PRIOR APPLICATION NUMBER: 09/295,612
;  PRIOR FILING DATE: 2000-04-20
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
;   LENGTH: 2662
;   TYPE: PRT
;   ORGANISM: Human
US-09-595-684B-31

  Query Match             7.0%;  Score 482;  DB 4;  Length 2662;
  Best Local Similarity   18.7%;  Pred. No. 1.6e-26;
  Matches  305;  Conservative  263;  Mismatches  607;  Indels  458;  Gaps   59;

Qy        5 IKVVVRKRPLSELEKK---------KKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEF 55
Db        7 VAVCVRVRPLNSREESLGETAQVYWKTDNNVI--------------YQVDGSK-----SF 47

Qy       56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD 115
Db       48 NFDRVFHGNETTKNVYEEIAAP-IIDSAIQGYNGTIFAYGQTASGKTYTMMGSE-----D 101

Qy      116 TPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLL---QKRKMVAALENGKK 172
Db      102 HLGVIPRAIHDIFQKIKKF-PDREFLLRVSYMEIYNETITDLLCGTQKMKPLIIREDVNR 160

Qy      173 EVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINK---- 228
Db      161 NVYVADLTEEVVYTSEMALKWITKGEKSRHYGETKMNQRSSRSHTIFRMILESREKGEPS 220

Qy      229 ----NTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDK--N 282
Db      221 NCEGSVKVSHLNLVDLAGSERAAQTGAAGVRLK-EGCNINRSLFILGQVIKKLSDGQVGG 279

Qy      283 HIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCIN 342
Db      280 FINYRDSKLTRILQNSLGGNPKTRIICTITPV--SFDETLTALQFASTAKYMKN-TPYVN 336

Qy      343 E--EDDTNTERI--SILDSK------GSEMNASSIEN----VVIKSNHLLSNNNNNKINR 388
Db      337 EVSTDEALLKRYRKEIMDLKKQLEEVSLETRAQAMEKDQLAQLLEEKDLLQKVQNEKIEN 396

Qy      389 ----------------------------GKINDKIERN--------------------N 399
Db      397 LTRMLVTSSSLTLQQELKAKRKRRVTWCLGKINKMKNSNYADQFNIPTNITTKTHKLSIN 456

Qy      400 ILK--NKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNN- 456
Db      457 LLREIDESVCSESDVFSNT------LDTLSEIEWNPATKLLNQENIESELNSLRADYDNL 510

Qy      457 ----------------NNNNNNDNNDNNNDNNNNNNNDSSSMVN------NMINHM---- 490
Db      511 VLDYEQLRTEKEEMELKLKEKNDLDEFEALERKTKKDQEMQLIHEISNLKNLVKHREVYN 570

Qy      491 --INNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDD----------- 537
Db      571 QDLENELSSKVELLREKEDQIKKLQEYIDSQKLENIKMDLSYSLESIEDPKQMKQTLFDA 630

Qy      538 MNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDP 597
Db      631 ETVALDAKRESAFLRSENLELKEKMKELAT---TYKQMENDIQLYQSQLEAKKKM-QVDL 686

Qy      598 QKNKDNTFFKSNINKM----EDNTPKDILYE-----------------SRNVSNMNGNV 635
Db      687 EKELQSAF--NEITKLTSLIDGKVPKDLLCNLELEGKITDLQKELNKEVEENEALREEVI 744

Qy      636 LLGLNKNTHHDI-----------------STKDE-----NHNDNKINNGVINIINNSNV 672
Db      745 LLSELKSLPSEVERLRKEIQDKSEELHIITSEKDKLFSEVVHKESRVQGLLEEIGKTKDD 804

Qy      673 NSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRY---VNEMDTSNKNNDNIFFDA 729
Db      805 LATTQSNYKSTDQEFQNFKTLH----------MDFEQKYKMVLEENERMNQEIVNLSKEA 854

Qy      730 ISCDNNM--------YPNITNNNNNNNNNNNNNNNNIDVENYNNRDG------------ 768
Db      855 QKFDSSLGALKTELSYKTQELQEKTREVQERLNEMEQLKEQLENRDSPLQTVEREKTLIT 914

Qy      769 -----TNNSMKLYAYNSHNL------FQPDNNKNTSNIQN-INTNKNNQDGNVNYSMNFC 816
Db      915 EKLQQTLEEVKTLTQEKDDLKQLQESLQIERDQLKSDIHDTVNMNIDTQE-----QLRNA 969

Qy      817 HYNLNDKNYLIDLNNK--EQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNM 874
Db      970 LESLKHQETINTLKSKISEEVSRNLH-MEENTGETKDEFQQKM---------------- 1011

Qy      875 GNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSNNNIY 934
Db     1012 ---------------VGIDKKQDLEAKNTQTLTADVKDNEIIEQQRKIFSLIQEKNELQ 1055

Qy      935 NNI---------------NDDDTFQN--------DYCHNDNTFTIRRKNNTNINSNIYQ 970
Db     1056 QMLESVIAEKEQLKTDLKENIEMTIENQEELRLLGDELKKQQEIVAQEKNHA-----IKK 1110

Qy      971 NDDIIYTINSLNDY-------------MSNTLLHFKEKYTYPTLSTNE------DIYNKE 1011
Db     1111 EGELSRTCDRLAEVEEKLKEKSQQLQEKQQQLLNVQEEMSEMQKKINEIENLKNELKNKE 1170

Qy     1012 MEGKHI---RLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDN---- 1064
Db     1171 LTLEHMETERLELAQKL---------NENYEEVKSITKERKVLKELQKSFETERDHLRGY 1221

Qy     1065 -NNVD----NDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMY------- 1112
Db     1222 IREIEATGLQTKEELKIAHIHLKEHQETIDELRRSVSEKTAQIINTQDLEKSHTKLQEEI 1281

Qy     1113 ---------------DDTKEILNNI-LLSKYKAEKDNVIKKYI-------------NE 1141
Db     1282 PVLHEEQELLPNVKKVSETQETMNELELLTEQSTTKDSTTLARIEMERLRLNEKFQESQE 1341

Qy     1142 DIKNMS-------------------LEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQI 1181
Db     1342 EIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQESQSKQEQSLNMKEKDNETTKI 1401

Qy     1182 NNETSD--------LRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESEN 1233
Db     1402 VSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDEMKSVAKEKDDLQRLQEVLQ--SESDQ 1459

Qy     1234 LRLLYQFLVVEY---------------QNKSANSVLLNVSSNNGDIILLNKKL--VQDN 1275
Db     1460 LKENIKEIVAKHLETEEELKVAHCCLKEQEETINELRVNLSEKETEISTIQKQLEAINDK 1519

Qy     1276 IKNSMDHNNIHKK 1288
```

```
Db       1520 LQNKI--QEIYEK 1530


RESULT 33
US-09-724-511-2
; Sequence 2, Application US/09724511
; Patent No. 6391601
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6391601el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1046
;  CURRENT APPLICATION NUMBER: US/09/724,511
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 09/632,344
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 367
;   TYPE: PRT
;   ORGANISM: Human
US-09-724-511-2

  Query Match              6.8%;  Score 465.5;  DB 4;  Length 367;
  Best Local Similarity   33.9%;  Pred. No. 1.6e-26;
  Matches  122;  Conservative   66;  Mismatches  137;  Indels   35;  Gaps    6;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVD-----------MTKYIER 52
                :||||| || :  ||         :: | :    |  |  : :|          : |  :
Db         12 MKVVVRVRPENTKEKAAGFHKVVHVVDKHILVFDPKQEEVSFFHGKKTTNQNVIKKQNKD 71

Qy         53 HEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYG 112
                :|: | |||:|        |:|:| || |:  : ||  |:  ||| ||:|||:|||||
Db         72 LKFVFDAVFDETSTQSEVFEHTTKP-ILRSFLNGYNCTVLAYGATGAGKTHTMLGS---- 126

Qy        113 QSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKK 172
                :| ||:         ::  ::   ::        :|: |:|  :: |||        :|  |: :|
Db        127 -ADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQK 185

Qy        173 EVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK------DI 226
                 |||  | : :  : ||::  : :|    |         |  ||||||:  | |:          |
Db        186 GVVVHGLTLHQPKSSEEILHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASI 245

Qy        227 NKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDK----- 281
              |:|  : |:: ||||||| | |           :| |||||||||  | |:    |
Db        246 NQNVRIAKMSLIDLAGSER-ASTSGAKGTRFVEGTNINRSLLALGNVINALADSKISFLC 304

Qy        282 -----NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKN 336
                   |||:|:|:||::|:|   |   ::||||  :||:      :  |  |||:|::|  |: |:
Db        305 FQRKNQHIPYRNSKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIKS 364


RESULT 34
US-09-723-097-2
; Sequence 2, Application US/09723097
; Patent No. 6492151
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6492151el motor proteins and methods for
;  TITLE OF INVENTION:  their.use
;  FILE REFERENCE: 1046
;  CURRENT APPLICATION NUMBER: US/09/723,097
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 09/632,344
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 367
;   TYPE: PRT
;   ORGANISM: Human
US-09-723-097-2

  Query Match              6.8%;  Score 465.5;  DB 4;  Length 367;
  Best Local Similarity   33.9%;  Pred. No. 1.6e-26;
  Matches  122;  Conservative   66;  Mismatches  137;  Indels   35;  Gaps    6;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVD-----------MTKYIER 52
                :||||| || :  ||         :: | :    |  |  : :|          : |  :
Db         12 MKVVVRVRPENTKEKAAGFHKVVHVVDKHILVFDPKQEEVSFFHGKKTTNQNVIKKQNKD 71

Qy         53 HEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYG 112
                :|: | |||:|        |:|:| || |:  : ||  |:  ||| ||:|||:|||||
Db         72 LKFVFDAVFDETSTQSEVFEHTTKP-ILRSFLNGYNCTVLAYGATGAGKTHTMLGS---- 126

Qy        113 QSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKK 172
                :| ||:         ::  ::   ::        :|: |:|  :: |||        :|  |: :|
Db        127 -ADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQK 185

Qy        173 EVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK------DI 226
                 |||  | : :  : ||::  : :|    |         |  ||||||:  | |:          |
Db        186 GVVVHGLTLHQPKSSEEILHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASI 245

Qy        227 NKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDK----- 281
              |:|  : |:: ||||||| | |           :| |||||||||  | |:    |
Db        246 NQNVRIAKMSLIDLAGSER-ASTSGAKGTRFVEGTNINRSLLALGNVINALADSKISFLC 304

Qy        282 -----NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKN 336
                   |||:|:|:||::|:|   |   ::||||  :||:      :  |  |||:|::|  |: |:
Db        305 FQRKNQHIPYRNSKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIKS 364


RESULT 35
US-09-632-344-2
; Sequence 2, Application US/09632344
; Patent No. 6534309
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6534309el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1046
;  CURRENT APPLICATION NUMBER: US/09/632,344
;  CURRENT FILING DATE:  2000-08-03
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 367
;   TYPE: PRT
;   ORGANISM: Human
US-09-632-344-2

  Query Match              6.8%;  Score 465.5;  DB 4;  Length 367;
  Best Local Similarity   33.9%;  Pred. No. 1.6e-26;
  Matches  122;  Conservative   66;  Mismatches  137;  Indels   35;  Gaps    6;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVD-----------MTKYIER 52
                :||||| || :  ||         :: | :    |  |  : :|          : |  :
Db         12 MKVVVRVRPENTKEKAAGFHKVVHVVDKHILVFDPKQEEVSFFHGKKTTNQNVIKKQNKD 71

Qy         53 HEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYG 112
                :|: | |||:|        |:|:| || |:  : ||  |:  ||| ||:|||:|||||
Db         72 LKFVFDAVFDETSTQSEVFEHTTKP-ILRSFLNGYNCTVLAYGATGAGKTHTMLGS---- 126

Qy        113 QSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKK 172
                :| ||:         ::  ::   ::        :|: |:|  :: |||        :|  |: :|
Db        127 -ADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQK 185

Qy        173 EVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK------DI 226
                 |||  | : :  : ||::  : :|    |         |  ||||||:  | |:          |
Db        186 GVVVHGLTLHQPKSSEEILHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASI 245

Qy        227 NKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDK----- 281
```

```
                |:|  : |:: ||||||| | |            :| ||||||||    | |:    |
Db        246 NQNVRIAKMSLIDLAGSER-ASTSGAKGTRFVEGTNINRSLLALGNVINALADSKISFLC 304

Qy        282 -----NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKN 336
                   |||:|:|:||::|:| .  |  ::|||| :||:     :  | ||:|::| |: |:
Db        305 FQRKNQHIPYRNSKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIKS 364

RESULT 36
US-09-541-782-4
; Sequence 4, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
;  APPLICANT:  Nislow, Corey
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  TITLE OF INVENTION:  Antifungal Assay
;  FILE REFERENCE:  1015
;  CURRENT APPLICATION NUMBER: US/09/541,782
;  CURRENT FILING DATE:  2000-04-03
;  NUMBER OF SEQ ID NOS:  10
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1038
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-541-782-4

  Query Match              6.8%;  Score 465;  DB 3;  Length 1038;
  Best Local Similarity   20.2%;  Pred. No. 7.5e-26;
  Matches  246;  Conservative  172;  Mismatches  366;  Indels  434;  Gaps   46;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKN---------NCTLYIDEPRYKVDMTKYIERHEF 55
              | | || |  :| |    | | :: | :            | |             :|  :   :
Db         75 ITVAVRCRGRNEREISMKSSVVVNVPDITGSKEISINTT-------GDTGITAQMNAKRY 127

Qy         56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPY--GQ 113
              ||||        :::      ||    |   |    |    ||  |  :||||||  |     |:
Db        128 TVDKVFGPGASQDLIFDEVAGPLFQD-FIKGYNCTVLVYGMTSTGKTYTMTGDEKLYNGE 186

Qy        114 -SDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYD--------------- 156
                || ||       :|   |  :    |    :   || |:|  :| |
Db        187 LSDAAGIIPRVLLKLFDTLELQQNDYV--VKCSFIELYNEELKDLLDSNSNGSSNTGFDG 244

Qy        157 ---------------------------------------------LLQKRKMVAALENGK 171
                                                           ||:||    :| |
Db        245 QFMKKLRIFASSTANNTTSNSASSSRSNSRNSSPRSLNDLTPKAALLRKRLRTKSLPNTI 304

Qy        172 KE--------------------------------------------VVVKDLKILRV 184
              |:                                               : :::|:   :
Db        305 KQQYQQQQAVNSRNNSSSNSGSTTNNASSNTNTNNGQRSSMAPNDQTNGIYIQNLQEFHI 364

Qy        185 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDL--KDINKNTSLGKIAFIDLAG 242
                    |  :  :   |:   |::        || |||||  |   |   |   ::     :  |:   :||||
Db        365 TNAMEGLNLLQKGLKHRQVASTKMNDFSSRSHTIFTITLYKKHQDELFRISKMNLVDLAG 424

Qy        243 SERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGK 302
              ||      :  :  | ::  :     | :||:|||  |      |     |||||:|:||::|:|   |
Db        425 SENINRSGALNQRAKEAG-SINQSLLTLGRVINALVDKSGHIPFRESKLTRLLQDSLGGN 483

Qy        303 SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEM 362
              :|: :|| |||      |:| :|| |:|: || |||                         ::
Db        484 TKTALIATISPAKVTSEETCSTLEYASKAKNIKNK---------------------PQL 521

Qy        363 NASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSS 422
               :  ::::::|              |  :    ||       ::::|  ||    :||    :      |  :
Db        522 GSFIMKDILVK---------NITMELAKI-----KSDLLSTKS----KEGIYMSQDHYKN 563

Qy        423 LN-DIDKIK------KNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNN 475
              || |::  :           |  :  :  |  :   |    |:    :    |     ::
Db        564 LNSDLESYKNEVQECKREIESLTSKNALLVKDKLKSKETIQSQNCQIES----------- 612

Qy        476 NNDSSSMVNNMINHM---INNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDF 532

                      :    |:|:     ::          | ::: ||                    |:
Db        613 -------LKTTIDHLRAQLDKQHKTEIEISDFNNKLQK--------------LTEVMQM 650

Qy        533 SSLDDMNCHLNNNDK---SIFLHKKNLRDNIKLK-NRSSCDNIMNKK--KNNLHLARHSV 586
              :  |          |:  |  |        |  |:   :  |:  |        ::|:  :     :  ||   :  ::  |
Db        651 ALHDYKKRELDLNQKFEMHITKEIKKLKSTLFLQLNTMQQESILQETNIQPNLDMIKNEV 710

Qy        587 GSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHD 646
                 ||:          :   |       :|    :   |:   :  :||
Db        711 ---LTLMR--TMQEKAELMYKDCVKKILNESPK------------------------- 738

Qy        647 ISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISD 706
                         .  .      |::                    :  |:    ::     ||    |    :::||
Db        739 ---------------FFNVV-----------IEKIDIIRVDFQKFYK---NIAENLSD 767

Qy        707 VQIRYVNEMDTSNKNND------NIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDV 760
                :             |  :||:           |  ||          |  :  :  |   |  :  ::      |          |   |
Db        768 I----------SEENNNMKQYLKNHFF------KNNHQELLNRHVDSTYENIEKRTNEFV 811

Qy        761 ENYNN--RDGTNNSMKLYAYN----------SHNLFQPDNNKNTSNIQNINTNKNNQDG 807
              ||:      |   :  :  ||   :|                   :||:|
Db        812 ENFKKVLNDHLDENKKLIMHNLTTATSAVIDQEMDLFEPK------------------- 851

Qy        808 NVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNI 867
               |  :  :|                   ||                |:  ||:              |:|                        :
Db        852 RVKWENSF-----------DL---------INDCDS----MNNEFYNSMAATLSQIKST 886

Qy        868 VIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSS 927
                ::|   | :        ||   :    |   :|    :|||     |::  :          |   |  :|  :|     ||  :| :
Db        887 VDTSSNSMNESISVMKGQVEESENAISLLKNNTKFNDQFEQLINKHNMLKDNIKNSITST 946

Qy        928 SSN----NNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLND 983
               |:    ::|||  |                                      ||    |    |   |   :        |:   |
Db        947 HSHITNVDDIYNTIE-----------------------NIMKN-YGNKE-----NATKD 976

Qy        984 YMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNV-- 1041
                 |   :|   ||      |  ||             |:|        :||                     :|:::|:   :|
Db        977 EMIENIL--KE---IPNLS-------KKMP---LRL-----------SNINSNSVQSVIS 1010

Qy       1042 ------DNNVDNNNVDN 1052
                      | |  :  ||||
Db       1011 PKKHAIEDENKSSENVDN 1028

RESULT 37
US-09-723-820-4
; Sequence 4, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
;  APPLICANT: Nislow, Corey
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  TITLE OF INVENTION: Antifungal Assay
;  FILE REFERENCE: 1015
;  CURRENT APPLICATION NUMBER: US/09/723,820
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: 09/541,782
;  PRIOR FILING DATE: 2000-04-03
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1038
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-723-820-4

  Query Match              6.8%;  Score 465;  DB 4;  Length 1038;
  Best Local Similarity   20.2%;  Pred. No. 7.5e-26;
  Matches  246;  Conservative  172;  Mismatches  366;  Indels  434;  Gaps   46;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKN---------NCTLYIDEPRYKVDMTKYIERHEF 55
              | | || |  :| |    | | :: | :            | |             :|  :   :
Db         75 ITVAVRCRGRNEREISMKSSVVVNVPDITGSKEISINTT-------GDTGITAQMNAKRY 127
```

```
Qy         56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPY--GQ 113
Db        128 TVDKVFGPGASQDLIFDEVAGPLFQD-FIKGYNCTVLVYGMTSTGKTYTMTGDEKLYNGE 186

Qy        114 -SDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYD--------------- 156
Db        187 LSDAAGIIPRVLLKLFDTLELQQNDYV--VKCSFIELYNEELKDLLDSNSNGSSNTGFDG 244

Qy        157 -------------------------------------------LLQKRKMVAALENGK 171
Db        245 QFMKKLRIFASSTANNTTSNSASSSRSNSRNSSPRSLNDLTPKAALLRKRLRTKSLPNTI 304

Qy        172 KE---------------------------------------------VVVKDLKILRV 184
Db        305 KQQYQQQQAVNSRNNSSSNSGSTTNNASSNTNTNNGQRSSMAPNDQTNGIYIQNLQEFHI 364

Qy        185 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDL--KDINKNTSLGKIAFIDLAG 242
Db        365 TNAMEGLNLLQKGLKHRQVASTKMNDFSSRSHTIFTITLYKKHQDELFRISKMNLVDLAG 424

Qy        243 SERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGK 302
Db        425 SENINRSGALNQRAKEAG-SINQSLLTLGRVINALVDKSGHIPFRESKLTRLLQDSLGGN 483

Qy        303 SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEM 362
Db        484 TKTALIATISPAKVTSEETCSTLEYASKAKNIKNK---------------------PQL 521

Qy        363 NASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSS 422
Db        522 GSFIMKDILVK---------NITMELAKI-----KSDLLSTKS----KEGIYMSQDHYKN 563

Qy        423 LN-DIDKIK------KNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNN 475
Db        564 LNSDLESYKNEVQECKREIESLTSKNALLVKDKLKSKETIQSQNCQIES----------- 612

Qy        476 NNDSSSMVNNMINHM---INNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDF 532
Db        613 -------LKTTIDHLRAQLDKQHKTEIEISDFNNKLQK---------------LTEVMQM 650

Qy        533 SSLDDMNCHLNNNDK---SIFLHKKNLRDNIKLK-NRSSCDNIMNKK--KNNLHLARHSV 586
Db        651 ALHDYKKRELDLNQKFEMHITKEIKKLKSTLFLQLNTMQQESILQETNIQPNLDMIKNEV 710

Qy        587 GSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHD 646
Db        711 ---LTLMR--TMQEKAELMYKDCVKKILNESPK-------------------------- 738

Qy        647 ISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISD 706
Db        739 ----------------FFNVV------------IEKIDIIRVDFQKFYK---NIAENLSD 767

Qy        707 VQIRYVNEMDTSNKNND------NIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDV 760
Db        768 I----------SEENNNMKQYLKNHFF------KNNHQELLNRHVDSTYENIEKRTNEFV 811

Qy        761 ENYNN--RDGTNNSMKLYAYN-----------SHNLFQPDNNKNTSNIQNINTNKNNQDG 807
Db        812 ENFKKVLNDHLDENKKLIMHNLTTATSAVIDQEMDLFEPK-------------------- 851

Qy        808 NVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNI 867
Db        852 RVKWENSF------------DL---------INDCDS----MNNEFYNSMAATLSQIKST 886

Qy        868 VIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSS 927
Db        887 VDTSSNSNMNESISVMKGQVEESENAISLLKNNTKFNDQFEQLINKHNMLKDNIKNSITST 946

Qy        928 SSN----NNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLND 983
Db        947 HSHITNVDDIYNTIE-----------------------NIMKN-YGNKE-----NATKD 976

Qy        984 YMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNV-- 1041
Db        977 EMIENIL--KE---IPNLS-------KKMP---LRL-----------SNINSNSVQSVIS 1010

Qy       1042 -------DNNVDNNNVDN 1052
Db       1011 PKKHAIEDENKSSENVDN 1028
```

RESULT 38 .
US-09-724-511-4
; Sequence 4, Application US/09724511
; Patent No. 6391601
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6391601el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1046
; CURRENT APPLICATION NUMBER: US/09/724,511
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 09/632,344
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Human
US-09-724-511-4

Query Match 6.8%; Score 464.5; DB 4; Length 355;
Best Local Similarity 34.0%; Pred. No. 1.8e-26;
Matches 122; Conservative 65; Mismatches 137; Indels 35; Gaps 6;

```
Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVD-----------MTKYIER 52
Db          4 MKVVVRVRPENTKEKAAGFHKVVHVVDKHILVFDPKQEEVSFFHGKKTTNQNVIKKQNKD 63

Qy         53 HEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYG 112
Db         64 LKFVFDAVFDETSTQSEVFEHTTKP-ILRSFLNGYNCTVLAYGATGAGKTHTMLGS---- 118

Qy        113 QSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKK 172
Db        119 -ADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQK 177

Qy        173 EVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK------DI 226
Db        178 GVVVHGLTLHQPKSSEEILHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASI 237

Qy        227 NKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDK----- 281
Db        238 NQNVRIAKMSLIDLAGSER-ASTSGAKGTRFVEGTNINRSLLALGNVINALADSKISFLC 296

Qy        282 -----NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFK 335
Db        297 FQRKNQHIPYRNSKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIK 355
```

RESULT 39
US-09-723-097-4
; Sequence 4, Application US/09723097
; Patent No. 6492151
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6492151el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1046
; CURRENT APPLICATION NUMBER: US/09/723,097
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 09/632,344

```
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 355
;   TYPE: PRT
;   ORGANISM: Human
US-09-723-097-4

  Query Match             6.8%;  Score 464.5;  DB 4;  Length 355;
  Best Local Similarity   34.0%;  Pred. No. 1.8e-26;
  Matches  122;  Conservative   65;  Mismatches  137;  Indels   35;  Gaps    6;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVD-----------MTKYIER 52
               :||||| || :  ||       :: | :    |  |  : :|          : |  :
Db          4 MKVVVRVRPENTKEKAAGFHKVVHVVDKHILVFDPKQEEVSFFHGKKTTNQNVIKKQNKD 63

Qy         53 HEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYG 112
                :|: | |||:|       |:|:| || |:  : ||  |:  ||| ||:|||:|||||
Db         64 LKFVFDAVFDETSTQSEVFEHTTKP-ILRSFLNGYNCTVLAYGATGAGKTHTMLGS---- 118

Qy        113 QSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKK 172
                :| ||:        ::  ::  ::       :|: |:|  :: |||      :|  |: :|
Db        119 -ADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQK 177

Qy        173 EVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK------DI 226
                |||  | : :  : ||::  : :|    |           |  ||||||:  | |:        |
Db        178 GVVVHGLTLHQPKSSEEILHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASI 237

Qy        227 NKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDK----- 281
               |:|  : |:: |||||||| | |            :| ||||||||    | |:    |
Db        238 NQNVRIAKMSLIDLAGSER-ASTSGAKGTRFVEGTNINRSLLALGNVINALADSKISFLC 296

Qy        282 -----NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFK 335
                      |||:|:|:||::|:|    |    ::|||| :||:      :  | |||:|::| |: |
Db        297 FQRKNQHIPYRNSKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIK 355
```

RESULT 40

```
US-09-632-344-4
; Sequence 4, Application US/09632344
; Patent No. 6534309
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6534309el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1046
;  CURRENT APPLICATION NUMBER: US/09/632,344
;  CURRENT FILING DATE:  2000-08-03
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 355
;   TYPE: PRT
;   ORGANISM: Human
US-09-632-344-4

  Query Match             6.8%;  Score 464.5;  DB 4;  Length 355;
  Best Local Similarity   34.0%;  Pred. No. 1.8e-26;
  Matches  122;  Conservative   65;  Mismatches  137;  Indels   35;  Gaps    6;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVD-----------MTKYIER 52
               :||||| || :  ||       :: | :    |  |  : :|          : |  :
Db          4 MKVVVRVRPENTKEKAAGFHKVVHVVDKHILVFDPKQEEVSFFHGKKTTNQNVIKKQNKD 63

Qy         53 HEFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYG 112
                :|: | |||:|       |:|:| || |:  : ||  |:  ||| ||:|||:|||||
Db         64 LKFVFDAVFDETSTQSEVFEHTTKP-ILRSFLNGYNCTVLAYGATGAGKTHTMLGS---- 118

Qy        113 QSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKK 172
                :| ||:        ::  ::  ::       :|: |:|  :: |||      :|  |: :|
Db        119 -ADEPGVMYLTMLHLYKCMDEIKEEKICSTAVSYLEVYNEQIRDLLVNSGPLAVREDTQK 177

Qy        173 EVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLK------DI 226
                |||  | : :  : ||::  : :|    |           |  ||||||:  | |:        |
Db        178 GVVVHGLTLHQPKSSEEILHLLDNGNKNRTQHPTDMNATSSRSHAVFQIYLRQQDKTASI 237

Qy        227 NKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDK----- 281
               |:|  : |:: |||||||| | |            :| ||||||||    | |:    |
Db        238 NQNVRIAKMSLIDLAGSER-ASTSGAKGTRFVEGTNINRSLLALGNVINALADSKISFLC 296

Qy        282 -----NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFK 335
                      |||:|:|:||::|:|    |    ::|||| :||:      :  | |||:|::| |: |
Db        297 FQRKNQHIPYRNSKLTRLLKDSLGGNCQTIMIAAVSPSSVFYDDTYNTLKYANRAKDIK 355
```

Search completed: October 2, 2003, 16:28:17
Job time : 44 secs

```
OM protein - protein search, using sw model

Run on:         October  2, 2003, 16:20:46 ; Search time 85 Seconds
                                             (without alignments)
                                             2397.392 Million cell updates/sec

Title:          US-10-006-780-2
Perfect score:  6842
Sequence:       1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters:      587654

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                17:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6842	100.0	1288	15	US-10-006-780-2	Sequence 2, Appli
2	1815	26.5	355	15	US-10-006-780-6	Sequence 6, Appli
3	1703	24.9	332	15	US-10-006-780-10	Sequence 10, Appl
4	1702	24.9	332	15	US-10-006-780-4	Sequence 4, Appli
5	1692	24.7	361	15	US-10-006-780-8	Sequence 8, Appli
6	768.5	11.2	363	15	US-10-159-151-10	Sequence 10, Appl
7	768.5	11.2	381	15	US-10-159-151-6	Sequence 6, Appli
8	768.5	11.2	1368	15	US-10-159-151-2	Sequence 2, Appli
9	758.5	11.1	343	15	US-10-159-151-8	Sequence 8, Appli
10	758.5	11.1	373	15	US-10-159-151-4	Sequence 4, Appli
11	750.5	11.0	725	12	US-09-849-602-20	Sequence 20, Appl
12	696.5	10.2	319	10	US-09-723-276-2	Sequence 2, Appli
13	696	10.2	807	11	US-09-820-843A-108	Sequence 108, App
14	694.5	10.2	861	11	US-09-820-843A-109	Sequence 109, App
15	676	9.9	2150	14	US-10-135-322-17	Sequence 17, Appl
16	672	9.8	410	10	US-09-925-300-1074	Sequence 1074, Ap
17	596	8.7	1331	15	US-10-087-464-50	Sequence 50, Appl
18	589.5	8.6	303	12	US-10-422-866-4	Sequence 4, Appli
19	589.5	8.6	492	12	US-10-422-866-2	Sequence 2, Appli
20	538.5	7.9	2184	12	US-10-304-095-6	Sequence 6, Appli
21	518.5	7.6	1610	15	US-10-155-533-9	Sequence 9, Appli
22	491	7.2	464	12	US-10-161-051-119	Sequence 119, App
23	460.5	6.7	1388	15	US-10-146-473-82	Sequence 82, Appl
24	454.5	6.6	1269	12	US-10-342-224-62	Sequence 62, Appl
25	453	6.6	383	10	US-09-883-096-5	Sequence 5, Appli
26	451	6.6	864	10	US-09-883-096-2	Sequence 2, Appli
27	446	6.5	1648	15	US-10-157-031-38	Sequence 38, Appl
28	429	6.3	338	10	US-09-883-096-4	Sequence 4, Appli
29	395.5	5.8	1103	10	US-09-847-874A-1	Sequence 1, Appli
30	388.5	5.7	596	12	US-10-167-831-2	Sequence 2, Appli
31	376.5	5.5	1253	12	US-10-363-798-2	Sequence 2, Appli
32	374.5	5.5	794	12	US-10-304-095-8	Sequence 8, Appli
33	373	5.5	1191	15	US-10-087-464-47	Sequence 47, Appl
34	348.5	5.1	665	11	US-09-820-843A-107	Sequence 107, App
35	342.5	5.0	6281	9	US-09-815-242-12996	Sequence 12996, A
36	336.5	4.9	460	10	US-09-925-300-1228	Sequence 1228, Ap
37	333.5	4.9	890	15	US-10-171-311-196	Sequence 196, App
38	333	4.9	652	9	US-09-351-794A-2	Sequence 2, Appli
39	327	4.8	1093	10	US-09-801-368-392	Sequence 392, App
40	325.5	4.8	959	12	US-10-342-224-40	Sequence 40, Appl
41	323.5	4.7	2086	9	US-09-815-242-5639	Sequence 5639, Ap
42	323.5	4.7	5795	9	US-09-815-242-12610	Sequence 12610, A
43	316.5	4.6	887	12	US-09-949-029-130	Sequence 130, App
44	314.5	4.6	1349	9	US-09-815-242-5898	Sequence 5898, Ap
45	314.5	4.6	1349	9	US-09-815-242-13137	Sequence 13137, A

ALIGNMENTS

```
RESULT 1
US-10-006-780-2
; Sequence 2, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
;  APPLICANT: Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
;  TITLE OF INVENTION:  METHODS FOR ITS USE
;  FILE REFERENCE: CYTOP083
;  CURRENT APPLICATION NUMBER: US/10/006,780
;  CURRENT FILING DATE:  2001-11-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1288
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-10-006-780-2

  Query Match             100.0%;  Score 6842;  DB 15;  Length 1288;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1288;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 60

Qy         61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 FDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF 120

Qy        121 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
```

```
Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240

Qy        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300

Qy        301 GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGS 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGS 360

Qy        361 EMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKY 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKY 420

Qy        421 SSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSS 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 SSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSS 480

Qy        481 SMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SMVNNMINHMINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 540

Qy        541 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600

Qy        601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 660

Qy        661 NGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNK 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 NGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNK 720

Qy        721 NNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNS 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 NNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNS 780

Qy        781 HNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIH 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 HNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIH 840

Qy        841 GCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNE 900

Qy        901 MKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKN 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 MKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKN 960

Qy        961 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLD 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLD 1020

Qy       1021 DQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIK 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 DQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIK 1080

Qy       1081 NFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYIN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 NFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYIN 1140

Qy       1141 EDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 EDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNN 1200

Qy       1201 NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSN 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSN 1260

Qy       1261 NGDIILLNKKLVQDNIKNSMDHNNIHKK 1288
              ||||||||||||||||||||||||||||
Db       1261 NGDIILLNKKLVQDNIKNSMDHNNIHKK 1288
```

```
RESULT 2
US-10-006-780-6
; Sequence 6, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
;  APPLICANT: Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
;  TITLE OF INVENTION:  METHODS FOR ITS USE
;  FILE REFERENCE: CYTOP083
;  CURRENT APPLICATION NUMBER: US/10/006,780
;  CURRENT FILING DATE:  2001-11-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 355
;   TYPE: PRT
;   ORGANISM: P. Falciparum
US-10-006-780-6

  Query Match             26.5%;  Score 1815;  DB 15;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-120;
  Matches  354;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 61

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 121

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 181

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 241

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 301

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDS 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDS 355
```

```
RESULT 3
US-10-006-780-10
; Sequence 10, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
;  APPLICANT: Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
;  TITLE OF INVENTION:  METHODS FOR ITS USE
;  FILE REFERENCE: CYTOP083
;  CURRENT APPLICATION NUMBER: US/10/006,780
;  CURRENT FILING DATE:  2001-11-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 332
```

```
;   TYPE: PRT
;   ORGANISM: P. Falciparum
US-10-006-780-10

  Query Match             24.9%;  Score 1703;  DB 15;  Length 332;
  Best Local Similarity   100.0%;  Pred. No. 4e-112;
  Matches  332;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 64
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 60

Qy         65 VDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAA 124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAA 120

Qy        125 GDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRV 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRV 180

Qy        185 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSE 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSE 240

Qy        245 RGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSK 304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 RGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSK 300

Qy        305 SIMIANISPTISCCEQTLNTLRYSSRVKNFKN 336
              ||||||||||||||||||||||||||||||||
Db        301 SIMIANISPTISCCEQTLNTLRYSSRVKNFKN 332

RESULT 4
US-10-006-780-4
; Sequence 4, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
;  APPLICANT: Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
;  TITLE OF INVENTION:  METHODS FOR ITS USE
;  FILE REFERENCE: CYTOP083
;  CURRENT APPLICATION NUMBER: US/10/006,780
;  CURRENT FILING DATE:  2001-11-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 332
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-10-006-780-4

  Query Match             24.9%;  Score 1702;  DB 15;  Length 332;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-112;
  Matches  332;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 60

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 120

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 180

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 240

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 300

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFK 335
              ||||||||||||||||||||||||||||||||
Db        301 KSIMIANISPTISCCEQTLNTLRYSSRVKNFK 332

RESULT 5
US-10-006-780-8
; Sequence 8, Application US/10006780
; Publication No. US20030104496A1
; GENERAL INFORMATION:
;  APPLICANT: Sakowicz, Roman
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: NOVEL MOTOR PROTEIN OF P. FALCIPARUM AND
;  TITLE OF INVENTION:  METHODS FOR ITS USE
;  FILE REFERENCE: CYTOP083
;  CURRENT APPLICATION NUMBER: US/10/006,780
;  CURRENT FILING DATE:  2001-11-30
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 361
;   TYPE: PRT
;   ORGANISM: P. Falciparum
US-10-006-780-8

  Query Match             24.7%;  Score 1692;  DB 15;  Length 361;
  Best Local Similarity   99.4%;  Pred. No. 2.7e-111;
  Matches  331;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 61

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 121

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 181

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 241

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 301

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKN 336
              |||||||||||||||||||||||||||||  |
Db        302 KSIMIANISPTISCCEQTLNTLRYSSRVKNKGN 334

RESULT 6
US-10-159-151-10
; Sequence 10, Application US/10159151
; Publication No. US20030036075A1
; GENERAL INFORMATION:
;  APPLICANT: Cytokinetics, Inc.
;  APPLICANT:  Beraud, Christopher
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  APPLICANT:  Patel, Umesh A.
;  APPLICANT:  Davies, Katherine A.
```

```
;  TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
;  FILE REFERENCE: 020552-002400US
;  CURRENT APPLICATION NUMBER: US/10/159,151
;  CURRENT FILING DATE:  2002-05-31
;  PRIOR APPLICATION NUMBER: US/09/967,908
;  PRIOR FILING DATE: 2001-09-28
;  PRIOR APPLICATION NUMBER: US 09/675,227
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 363
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Peptide
;   OTHER INFORMATION: fragment E213 - S566.
US-10-159-151-10

  Query Match             11.2%;  Score 768.5;  DB 15;  Length 363;
  Best Local Similarity   47.5%;  Pred. No. 3.2e-46;
  Matches  164;  Conservative   58;  Mismatches  116;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| |||||    | :: : :||||::  || : | :  ||:|:|| :| |  |:|| :
Db         12 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 71

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                 |  ||  |  |||  :: ||   :||||||||:||||||:|:      : ||:: |
Db         72 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 125

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
               | |||  | :          ::|||||||||:||||| :|| : | |: |   | :  |: |:
Db        126 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 185

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
               | : | |:  :: |    |  |    | :||||||:: | :||  | |  |:|:|||||||
Db        186 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 244

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              || ||     ::||: :|| ||:||||||||||:| :   | ||| |:||:||:| |:| :
Db        245 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 304

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
              |: ||||||:    | |||||||: |||  |      |
Db        305 KTCMIANISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRN 349
```

RESULT 7
US-10-159-151-6
; Sequence 6, Application US/10159151
; Publication No. US20030036075A1
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christopher
; APPLICANT: Guo, Jun
; APPLICANT: Freedman, Richard
; APPLICANT: Patel, Umesh A.
; APPLICANT: Davies, Katherine A.
; TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
; FILE REFERENCE: 020552-002400US
; CURRENT APPLICATION NUMBER: US/10/159,151
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US/09/967,908
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/675,227
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 381
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Peptide
;   OTHER INFORMATION: fragment V195-S566 with flanking vector sequence.
US-10-159-151-6

```
  Query Match             11.2%;  Score 768.5;  DB 15;  Length 381;
  Best Local Similarity   47.5%;  Pred. No. 3.4e-46;
  Matches  164;  Conservative   58;  Mismatches  116;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| |||||    | :: : :||||::  || : | :  ||:|:|| :| |  |:|| :
Db         30 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 89

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                 |  ||  |  |||  :: ||   :||||||||:||||||:|:      : ||:: |
Db         90 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 143

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
               | |||  | :          ::|||||||||:||||| :|| : | |: |   | :  |: |:
Db        144 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 203

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
               | : | |:  :: |    |  |    | :||||||:: | :||  | |  |:|:|||||||
Db        204 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 262

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              || ||     ::||: :|| ||:||||||||||:| :   | ||| |:||:||:| |:| :
Db        263 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 322

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
              |: ||||||:    | |||||||: |||  |      |
Db        323 KTCMIANISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRN 367
```

RESULT 8
US-10-159-151-2
; Sequence 2, Application US/10159151
; Publication No. US20030036075A1
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christopher
; APPLICANT: Guo, Jun
; APPLICANT: Freedman, Richard
; APPLICANT: Patel, Umesh A.
; APPLICANT: Davies, Katherine A.
; TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
; FILE REFERENCE: 020552-002400US
; CURRENT APPLICATION NUMBER: US/10/159,151
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US/09/967,908
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/675,227
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1368
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: KinI-3 motor
;   OTHER INFORMATION: domain fragment
US-10-159-151-2

```
  Query Match             11.2%;  Score 768.5;  DB 15;  Length 1368;
  Best Local Similarity   47.5%;  Pred. No. 1.9e-45;
  Matches  164;  Conservative   58;  Mismatches  116;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| |||||    | :: : :||||::  || : | :  ||:|:|| :| |  |:|| :
Db        223 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 282

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
```

```
               |   ||   |   |||  :: ||    :||||||||:||||||:|:        : ||:: |
Db         283 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 336

Qy         124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
               | |||  | :           ::||||||||||:|||||  :||  :  |  |:  |   |  :   |: |:
Db         337 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 396

Qy         184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
               | : | |:  :: |      |    |       | :||||||:: | :||   | |   |:|:|||||||
Db         397 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 455

Qy         244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
               || ||     ::||: :||  ||:||||||||||||:|  :     | |||  |:||:||:|  |:| :
Db         456 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 515

Qy         304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
               |: |||||||:      | |||||||: |||   |      |         |
Db         516 KTCMIANISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRN 560


RESULT 9
US-10-159-151-8
; Sequence 8, Application US/10159151
; Publication No. US20030036075A1
; GENERAL INFORMATION:
;  APPLICANT: Cytokinetics, Inc.
;  APPLICANT:  Beraud, Christopher
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  APPLICANT:  Patel, Umesh A.
;  APPLICANT:  Davies, Katherine A.
;  TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
;  FILE REFERENCE: 020552-002400US
;  CURRENT APPLICATION NUMBER: US/10/159,151
;  CURRENT FILING DATE:  2002-05-31
;  PRIOR APPLICATION NUMBER: US/09/967,908
;  PRIOR FILING DATE: 2001-09-28
;  PRIOR APPLICATION NUMBER: US 09/675,227
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 343
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Peptide
;   OTHER INFORMATION: fragment E213-S546.
US-10-159-151-8

  Query Match             11.1%;  Score 758.5;  DB 15;  Length 343;
  Best Local Similarity   48.9%;  Pred. No. 1.5e-45;
  Matches  161;  Conservative   58;  Mismatches  103;  Indels    7;  Gaps    3;

Qy           4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
               ||:| ||||||   | :: : :||||::  || :  | :   ||:|:|| :|  |   |:|| :
Db          12 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 71

Qy          64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                    |   ||   |   |||  :: ||    :||||||||:||||||:|:        : ||:: |
Db          72 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 125

Qy         124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
               | |||  | :           ::||||||||||:|||||  :||  :  |  |:  |   |  :   |: |:
Db         126 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 185

Qy         184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
               | : | |:  :: |      |    |       | :||||||:: | :||   | |   |:|:|||||||
Db         186 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 244

Qy         244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
               || ||     ::||: :||  ||:||||||||||||:|  :     | |||  |:||:||:|  |:| :
Db         245 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 304
```

```
Qy         304 KSIMIANISPTISCCEQTLNTLRYSSRVK 332
               |: |||||||:      | |||||||: |||
Db         305 KTCMIANISPSHVATEHTLNTLRYADRVK 333

RESULT 10
US-10-159-151-4
; Sequence 4, Application US/10159151
; Publication No. US20030036075A1
; GENERAL INFORMATION:
;  APPLICANT: Cytokinetics, Inc.
;  APPLICANT:  Beraud, Christopher
;  APPLICANT:  Guo, Jun
;  APPLICANT:  Freedman, Richard
;  APPLICANT:  Patel, Umesh A.
;  APPLICANT:  Davies, Katherine A.
;  TITLE OF INVENTION: KINI-3 MOTOR PROTEIN AND METHODS FOR ITS USE
;  FILE REFERENCE: 020552-002400US
;  CURRENT APPLICATION NUMBER: US/10/159,151
;  CURRENT FILING DATE:  2002-05-31
;  PRIOR APPLICATION NUMBER: US/09/967,908
;  PRIOR FILING DATE: 2001-09-28
;  PRIOR APPLICATION NUMBER: US 09/675,227
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 373
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Fragment
;   OTHER INFORMATION: encoding residues D183-L546 with flanking vector
;   OTHER INFORMATION: sequence.
US-10-159-151-4

  Query Match             11.1%;  Score 758.5;  DB 15;  Length 373;
  Best Local Similarity   48.9%;  Pred. No. 1.7e-45;
  Matches  161;  Conservative   58;  Mismatches  103;  Indels    7;  Gaps    3;

Qy           4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
               ||:| ||||||   | :: : :||||::  || :  | :   ||:|:|| :|  |   |:|| :
Db          42 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 101

Qy          64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                    |   ||   |   |||  :: ||    :||||||||:||||||:|:        : ||:: |
Db         102 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 155

Qy         124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
               | |||  | :           ::||||||||||:|||||  :||  :  |  |:  |   |  :   |: |:
Db         156 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 215

Qy         184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
               | : | |:  :: |      |    |       | :||||||:: | :||   | |   |:|:|||||||
Db         216 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 274

Qy         244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
               || ||     ::||: :||  ||:||||||||||||:|  :     | |||  |:||:||:|  |:| :
Db         275 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 334

Qy         304 KSIMIANISPTISCCEQTLNTLRYSSRVK 332
               |: |||||||:      | |||||||: |||
Db         335 KTCMIANISPSHVATEHTLNTLRYADRVK 363

RESULT 11
US-09-849-602-20
; Sequence 20, Application US/09849602
; Publication No. US20030165834A1
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Matthew J.
;  APPLICANT:  Old, Lloyd J.
```

```
;  APPLICANT:  Stockert, Elisabeth
;  APPLICANT:  Chen, Yao-Tseng
;  TITLE OF INVENTION: Colon Cancer Antigen Panel
;  FILE REFERENCE: L0461/7105(JRV)
;  CURRENT APPLICATION NUMBER: US/09/849,602
;  CURRENT FILING DATE:  2001-05-04
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
;   LENGTH: 725
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-849-602-20

  Query Match             11.0%;  Score 750.5;  DB 12;  Length 725;
  Best Local Similarity   41.9%;  Pred. No. 1.5e-44;
  Matches  173;  Conservative   76;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                :| | |||||:: |  ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db        258 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 317

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
                |  |  ||   | :||:  ::| |   :|||||||||||:|| |        |: : ||:
Db        318 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 376

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
                |: |:|    |    | |     : :::|:|||  |||:|||  |:  :    ||:||::| |  |:
Db        377 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 435

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                 | : :: ::||:|  |      |       ||||||   ||:    |          || : :|
Db        436 EHLVNSADD-VIKMLDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 492

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
                |||:||||||| | ::|:|: :|| ||:|||||||||||:   :| | |||:|:||:|||| |
Db        493 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 552

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
                :|: |:: ||| ||| || || ||||||: |||       |       |:              :: :
Db        553 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 603

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                  || | |   |  :     :|               |: : :     |: |: :: :|
Db        604 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 654

RESULT 12
US-09-723-276-2
; Sequence 2, Application US/09723276
; Patent No. US20020101952A1
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. US20020101952A1el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1032
;  CURRENT APPLICATION NUMBER: US/09/723,276
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 09/675,227
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 319
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (154)...(183)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-723-276-2

  Query Match             10.2%;  Score 696.5;  DB 10;  Length 319;
  Best Local Similarity   45.8%;  Pred. No. 3.2e-41;
  Matches  149;  Conservative   51;  Mismatches  118;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                ||:| |||||     | :: : :|||::  || :  | :  ||:|:|| :| |  |:|| :
Db          1 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 60

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                  |  ||  |  |||  :: ||   :|||||||:||||||:|:         : ||::  |
Db         61 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 114

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
                 | |||  | :            ::|||||||||:||||| :||
Db        115 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKXXXXXXXXXXXXXXXXXXXXX 174

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
                    . :: |   |     | :||||||:: | :||  | |  |:|:|||||||
Db        175 XXXXXXXXXVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 233

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
                || ||       ::|| : :|| ||:||||||||||:|  :  | ||| |:|:||:|: |:| :
Db        234 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 293

Qy        304 KSIMIANISPTISCCEQTLNTLRYS 328
                |: |||||||:     | |||||||:
Db        294 KTCMIANISPSHVATEHTLNTLRYA 318

RESULT 13
US-09-820-843A-108
; Sequence 108, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
;  APPLICANT: Council of Scientific and Industrial Research
;  TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTE
;  TITLE OF INVENTION:  USEFUL AS ANTI-INFECTIVES
;  FILE REFERENCE: Q63915
;  CURRENT APPLICATION NUMBER: US/09/820,843A
;  CURRENT FILING DATE:  2001-03-30
;  NUMBER OF SEQ ID NOS: 118
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
;   LENGTH: 807
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: hypothetical protein
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: gi|3845292
US-09-820-843A-108

  Query Match             10.2%;  Score 696;  DB 11;  Length 807;
  Best Local Similarity   26.6%;  Pred. No. 1.2e-40;
  Matches  253;  Conservative  151;  Mismatches  300;  Indels  246;  Gaps   43;

Qy        328 SSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKIN 387
                ||: || :|     ||: ||:|           | | |  ::| | |   :  :||:| || |
Db         10 SSKEKNEEN--DIINKCDDSN-------KINGKE-NIFAVEKVGINESGHMSNDNINK-N 58

Qy        388 RGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYN- 446
                 : |    | :: |  |  : :            :|  | :    ||:|  :  :: :|| |:  |
Db         59 QEKNKKKKKKNTHKKVNINNTHINIHTTNDKNNG-QDINKPEVIERDNIINIKNDTNNI 117

Qy        447 -DNTINKKHNNNNNNNDNNDNNNDNNNNNNDSSSMVNN-----MINHMINNNI----- 495
                 |:: |:: | || |: :||:|||: | |||| ::|  ||        |   :   |:|
Db        118 LDSSYNEEGNENNRNDINNNNNNNNININNNNINNSCSNNYGLKKKITLLKRNDIKDEGY 177

Qy        496 -NNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKK 554
                 | ||   || ||  |||| ::|                          |||:
Db        178 NNENITTLNNKNNLKNNNNYNDNR-----------------------NNNN-------- 205
```

```
Qy    555 NLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKME 614
Db    206 NNKNNINNNNNNNC-----------------CSEKTLEQREKEYNKIRARIFSNFNKKQ 247

Qy    615 DNTPKDILYESRNVSN--MNGNVLLGLNKNTHHDISTKDENHNDNK---INNGVINIINN 669
Db    248 KNVQKT---EQNNLNHTYLNNNIINNI-------------NNGDNQYAYINN-FYHIYHN 290

Qy    670 SNVNSINNSNMNSINNSNMNSNSIYKSN----YNSNQSISDVQIRYVNEM-DTSNKNNDN 724
Db    291 NSYNHIYRQNNIPICNINNHAPNIEKLNNPYYYHDN------HIAYTNYMYSTQNKM--- 341

Qy    725 IFFDAISCDNNM---------YPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKL 775
Db    342 ---------NNMKTKQIGHYGINNEDNNNNNNNNINNNNNNNNINNNNINN---NNVPLCI 389

Qy    776 YAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLID-LNNKE- 833
Db    390 PQLDNYNKTKNNFNQGTNNFNQGTNNFNKCTNNFNNAKNHIKHNINNTNKNIEHLNNHSI 449

Qy    834 -----QKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCG 888
Db    450 YNFVYPENKNIYDANGNLINNNISYTQLKMNNNINFNIHMESPINQQHNNTFKV-----N 504

Qy    889 SHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSNNN--IYNNINDDDTFQND 946
Db    505 NDTNFFNEPTNKMKK---KNKEKKNIHFNNNNNN-------NNNKCLYKDIN-----QND 549

Qy    947 YCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNED 1006
Db    550 H-----------NNSIINTN--QNFDHINNV-------KNTEQNLQKKHNKMSQVSKQS 588

Qy   1007 IYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNV---DNNNVDNNVDN------- 1056
Db    589 NNKNNKNNSHLK----KQININTNNNMDNKNNSHISKNVIVDDNKLKSSHADNSNEIVTK 644

Qy   1057 ----------------NDKNNVDN----NNVDNDDDDVDFHNIKNFNNNEY-----LSYF 1091
Db    645 GKKKKNTNKKKKINNINSVNNVNNINSMNNINSMNNIISMNNVNNMNNPMYFPNVNIQKD 704

Qy   1092 QKNVDTIINNCLNSLDISSM---YDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSL 1148
Db    705 DSNIALLYNNKPN-IDFNNFQLNHINNHMIQNNIMTNNVML-NNNLTTSNFNYNLINYSY 762

Qy   1149 E---------EIDKTAQ-SIYEKRKVLLTKLLLLFKKNVDTQINNETSDL 1188
Db    763 EPFYEENLMNDLDYCRDISLYEKR----------YDRGDNLQQNHKRYDI 802
```

RESULT 14
US-09-820-843A-109
; Sequence 109, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 109
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc_feature
; OTHER INFORMATION: gi|4493994
US-09-820-843A-109

```
  Query Match             10.2%;  Score 694.5;  DB 11;  Length 861;
  Best Local Similarity   26.4%;  Pred. No. 1.6e-40;
  Matches  254;  Conservative  145;  Mismatches  325;  Indels  237;  Gaps   41;

Qy    393 DKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINK 452
Db     21 DILNKYCFIRERKYNKPGG------NKY-----IPRDRSNNNNNIGNNVNGMNNFVLLNN 69

Qy    453 KHNNNNNNNNDNNDNNNDNNNNNNNDSSMVNNMINHMINNNINNNI-NVNNNNNNNNNN 511
Db     70 NNNNMRIRNTYNNNNNNINNNNNNN------NNNFNNFNNNNNNNFNNFNNFNNNNNFN 123

Qy    512 NNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSC--- 568
Db    124 NNNHFNIHNIDNYDDSYVKGRHRGNYLSSSLNNINGKVF---KNLDDN--------CYNL 172

Qy    569 --DNIMNKKKNNLHLA-----------RHSVGSKLTMFSYDPQKNKDNTFFKSNINKMED 615
Db    173 PTNNLYIDKEGKMHLTGKEHYNAASSNEYNHNNKNTN-NYNNNSYNNNNFCNNNYNDNNY 231

Qy    616 NTPKD-------------------ILYESRNVSNMN--------GNVLLGLNKNTHHDI 647
Db    232 NNSNNKGMGNKYERSLNYLKKEHDMVDYEYNNKGNIRKNDSEKYWDNPPLHYSKKNNYDI 291

Qy    648 ST-----KDENHNDNKINNGVINI----INNSNVNSINNSNMNSINNSNMNSNSIYKSNY 698
Db    292 FTLGDIKKYAKNNEKKGNNKYMNMHDNNSNNSN-NVLNNNNMNS--NSN-NYNNIFKDND 347

Qy    699 NSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNM----YPNITNNNNNNNNN--NN 752
Db    348 EENLTKSN----FAKWFKNNNNMNVNENTDIIKYLNNKNSQGHSDGKNNNNNNGNNIINN 403

Qy    753 NNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYS 812
Db    404 NSNNKNNIFQGNSRNYENVMYNINNNNNNNIISNNKNEASFNTDNINTNSGREEEKIS-- 461

Qy    813 MNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRND----FEKKKKTNFYNNNNIV 868
Db    462 -----------NTVAELLMKQ----------ISMIKERNKGLDVLEKKNTFGFLDNN--- 497

Qy    869 IVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSS 928
Db    498 --YQNYGSNNN---------SSLEKNNMKENDIYSKEASKRIM-DIFRTLNSNGLVSQES 545

Qy    929 --------SNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINS 980
Db    546 LLVNQSVLNNNNNYNNYNSNNNRNKNQNNNNN-------NNNNMNMNMNSNNI---NNN 595

Qy    981 LNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNN 1040
Db    596 NNYYKNNHKYHSMDNVTYKKIFIN-------------------NYSNNDGNN--NSNNSN 634

Qy   1041 VDNNVD----NN--NVDNNVDN-----NDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLS 1089
Db    635 SNNNVEHYYMNNKKNFKNKINNYHNLPDNKNNMMNNNTYNNINKNNLSNMENFPPS--LS 692

Qy   1090 YFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSLE 1149
Db    693 F--NNSDINKNNAQGNINIT-------PIINSIL--RLDNEVDNVHNNSISENIQNAKVS 741

Qy   1150 EIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFY 1209
Db    742 NVLDSLKSLLKASKSQGNN-----NYNIPKNFNN---------------NNNNNNNSKFI 781

Qy   1210 AYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNNGDIILLNK 1269
Db    782 NYNS----------QQYYPSHQQQQQHQ----QQQQQQQQQTLIQTQINSTHLNDFNK 826

Qy   1270 K 1270
Db    827 K 827
```

```
RESULT 15
US-10-135-322-17
; Sequence 17, Application US/10135322
; Publication No. US20020173017A1
; GENERAL INFORMATION:
;  APPLICANT: BENFEY, PN
;  APPLICANT:  HELARIUTTA, Y
;  APPLICANT:  MAHONEN, AP
;  APPLICANT:  BONKE, AWM
;  APPLICANT:  KAUPPINEN, L
;  APPLICANT:  RIIKONEN, M
;  TITLE OF INVENTION: WOODEN LEG GENE, PROMOTER AND USES THEREOF
;  FILE REFERENCE: 5914-086-999
;  CURRENT APPLICATION NUMBER: US/10/135,322
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: 60/253,739
;  PRIOR FILING DATE: 2000-11-29
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
;   LENGTH: 2150
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-135-322-17

  Query Match              9.9%;  Score 676;  DB 14;  Length 2150;
  Best Local Similarity   24.5%;  Pred. No. 1.1e-38;
  Matches  242;  Conservative  146;  Mismatches  255;  Indels  344;  Gaps   41;

Qy        416 TFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNN 475
              ::| :| ::::                :|      |:|  |   :||||||||:|| |||:|||||
Db         85 SYGNHSFIHNVSPTSP-------SYDINNNNNNNNNNNNNNNNNNNNNNSNNNNNNNNN 137

Qy        476 N------------------------------------------NDSSSM--VNNMINH 489
              |                                          |:|||:  :|  :|
Db        138 NKNNYNNNYYYSPIENSNISKSLEESVLNQFPHNFNLSSNNNYLNNSSSLHNINQSVNS 197

Qy        490 MINNNINNN----INVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNN 545
              : ||| |       || ||||||||||||:|::                    | : |||
Db        198 LSNNNNNQTNQQPINNNNNNNNNNNNNNSNNSN-------------------NSNNNNN 237

Qy        546 DKSIFLHKKNLRDN-IKLKNRSSCDNIMNKKKNNLHLARHSVGSKL---TMFSYDPQKNK 601
                : |: |:  | |  |: :  :           |  | |: | :    : |:   |||
Db        238 GN----NNNNITDSPTKSKRHSTYETNIGS-----HQRRKSIQSLIANSAIHSFSKLKNK 288

Qy        602 DNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINN 661
                          :  :||   :              |      :|  |:::       |:|:|   :
Db        289 ----------PLSSSTPSTV-----------NTCGAVNNNSNN-------NNNNNNNST 319

Qy        662 GVINIIN-----NSNVNSI-------NNS-------------------NMNSINNS-NM 688
              | :  |        : |:|:|          |||                   |::|:|:  |:
Db        320 GSLGAIPMDRSFDGNINTITEESTGGNNSPRSNCGSNCGSNGGIPLSPRNLSSLNSGVNV 379

Qy        689 NSNSIYKSNYNSNQS-ISDVQIRYVN-EMDTSNKNNDNIFFDAISCDNNMYPNITNNNNN 746
              :  :|: :| |:|  |  :  :  |::|  :: || ||:|            :||: ||   ||:|
Db        380 SPRNIHLNNLNNNSSNLPPLSPRHINFHINVSNLNNNN--------NNNINPNNNPNNSN 431

Qy        747 NNNNN---NNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKN 803
              |:|||    |||:||      |       :||                  ||  | :|| |
Db        432 NSNNNVSPRNNNHNISPRGSNISPRSNN-------GGSTTISPRNISNNNNIIN------ 478

Qy        804 NQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCD---------NNIIQNRNDFE 854
                              |:|: | |   |  : : |::  :          |: :   :
Db        479 ---------------NINNNNILTPPRNSPRLE-NVNPTNSPRLLATSLNSTLPIVSSLT 522

Qy        855 KKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDN 914
                  |  |||    :|||  ||           ||     |  :  :  |  :    :|
Db        523 SSNNNNQSNNNTNPSINNNNGRN-------GHCIQTISEEILGNKPVVYNNGNNN----- 570

Qy        915 HIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQ---- 970
                  ||||:||::|:||||  |  |::      .              || |||:|:
Db        571 ----NNNNTNNSTTSNNNITTNNNNN-------------------NNNNINNNVLSTPRK 607

Qy        971 ----NDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYD 1026
                  |      ||| |: ::                          ::: :|  :    |  :|  ::|  |
Db        608 RTKGNHS---KTNSLQDFETS------------SMNGGDDSISGAGSGGSLRRRNKDDND 652

Qy       1027 DND-------------NNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDD 1073
              :||              ||| :||: || :|| :||| :|| :||: || :||| :|::::
Db        653 ENDGNSNNTNSNNSNNNNNNNNNSSNNNNNNSNNNNNNNNNNNNNNNNNNNNNNNNNNNN 712

Qy       1074 VDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDN 1133
                : :|  |||                     ||  |     :     :          |:   :     :| |
Db        713 NNNNNNNNNNNNN-------------NNYHNGATMMMSH-------NHQSIGMSSSPKKN 752

Qy       1134 VIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLV 1193
                 | :     :| |·|  : : | :|              :| ||          |    :   :   ||
Db        753 NFKPF----SRNCSLMGMGRRAWAI----------ILGLF-------IVGSSISILATLV 791

Qy       1194 MCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNK----S 1249
                : :     |:  | |.      :|: :|      |||                     :|: |:
Db        792 LRYSEENSIADDF-----ARVARD--RFTMLR-----------------IEFNNRLYITQ 827

Qy       1250 ANSVLLNV-SSNNGDIILLNKKLVQDN 1275
                 |:||:|  | : |  :    || ||
Db        828 TLSLLLSVFPSTSEDQFVPFSKLWSDN 854


RESULT 16
US-09-925-300-1074
; Sequence 1074, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1074
;   LENGTH: 410
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (14)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (21)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (177)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (248)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (300)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (372)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1074

  Query Match              9.8%;  Score 672;  DB 10;  Length 410;
  Best Local Similarity   46.6%;  Pred. No. 2.4e-39;
  Matches  149;  Conservative   59;  Mismatches  102;  Indels   10;  Gaps    8;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | |||||::| ||: |:|:: : | | : ||: |||:|||:|   | | ||:
```

```
Db         57 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 116

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |   |  ||   | :||:  ::| |    :|||||||||||:|| |       |: : ||:
Db        117 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 175

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              | |:|   |    | |    : ::::|:||| |||:||| |:  :  ||:||::| |  |:
Db        176 AXRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 234

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                | : :: ::|| |  |    |      ||||||    | |:    |     || : :|
Db        235 EHLVNSADD-VIKMXDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 291

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:|||| |  | ::||: :|| ||:||||||||||:   :| |  |||:|:||:|||| |
Db        292 LAGNERGAXTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 351

Qy        300 VGK-SKSIMIANISPTISCC 318
              :|: |:: ||| ||| || |
Db        352 IGENSRTCMIATISPGISSC 371
```

RESULT 17
US-10-087-464-50
; Sequence 50, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 50
;   LENGTH: 1331
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-10-087-464-50

```
  Query Match             8.7%;  Score 596;  DB 15;  Length 1331;
  Best Local Similarity   22.3%;  Pred. No. 2.6e-33;
  Matches  317;  Conservative  180;  Mismatches  412;  Indels  512;  Gaps   68;

Qy        122 YAAGDIFTFLNIYDKDNTK------GIFISFY---EIYCGKLYDLLQKRKMVAA------ 166
               | :| |:  ||  |   |           :  ||     |:| |   ||:     :| |
Db         52 YVSSDLLNFLKCYSNLNINLNKVPYDLVYSFLLDGELYLG--YDISVFILLVKAEHFEYC 109

Qy        167 ----LENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNID 222
                   || |:   :          ||: :  :|            :|| |:       :| |
Db        110 RRIDNENSDKKESFR--------TKNKSTIKR-----------SSQIDDEDNLQGLLIKE 150

Qy        223 LKD----INKNTSLGKIAFIDLAGSERGADT---VSQNKQTQTDGANINRSLLALKECIR 275
               :|     :|:|      |       ||: :   :  : ::|    |    |    :|  |
Db        151 KEDYLSFLNENNE----ALKQYMESEKRGNPLWHLDESKYMDKDWYDEEDSSFIFKPTFN 206

Qy        276 AM-------DSDKNHIPFRDSELTKVLRD-----IFVGKSKSIMIANISPTISCCEQTLN 323
                :        :: |:  | :  :  :  |        || |  | : | :        | :|:
Db        207 YLGKNNNNNNNHNNNNAFSNFVMGNLSSDNISGCFFVEKLNAYLFAMLD---KCSNKTVI 263

Qy        324 TL-------RYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNH 376
               ::          |: ||      | :     :: :|       |           : |  |: |: ||
Db        264 SVFPYEKFGRHESR-----NLAIQFSQYEDYMHRIIEDRLYANIQNNLPSVHNMKNMSN- 317

Qy        377 LLSN-NNNNK---INRGKINDKIERN-----NILKNKSFDKPREGFTSTFGKYSSLNDID 427
               ::| |||||    |||  |::    ::      |||      :|| :         | :|   :|
Db        318 -MNNINNNNKDIIINRSGISNGNSQSVPCFENIL---DYDKLK------FVEY--INSFS 365

Qy        428 KIKKNKKKGLINYKSTLYNDNTINKK----HNNNNNNNNDNN----------DNNNDNN 472
               :||:    :|     :|      |     |||  :   :|              :|: |
Db        366 DVKKSSSFDIIGSSKNIYEQGENLKNYCIYHNNNFESGFENYILENKQPLELIENHFDIM 425

Qy        473 NN------NNNDSSSMVNNMINHMINNNIN-NNINVNNNNNNNNNNNNSHNNHLPQPNYA 525
               |        | |        ||:    :    :| |  |  |   :   :|:||| :  |  |  :      |
Db        426 ENIKGMYDNTNQEEMNFNNVSGLLREDNSNMNEIYLTRDNHNNNYHENEENIYSINIKY- 484

Qy        526 FTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHS 585
                           :| | |  |   |     ||:: :| : ||  |||       ||  :|  |:
Db        485 ------------INNHFNNKDDMI-MKCKNMKGSISMDNNSS-----NSNSNNTHFEK-- 524

Qy        586 VGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHH 645
                       |: |  :|       |:     |  | :: |:  |        |        :       |:|   : |:|:
Db        525 -----TLESINPD---DHNIFNSEMDSMK-NENND---EEEQTATSIYNILGKIGKDTY- 571

Qy        646 DISTKDENHN-DNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSI 704
               |     |:| ||  |||     | |:  |    |    |  :: ||:   |||      |:|||
Db        572 -IKRCSSNYNYDN--NNG---YSNESSDNYNNGYNDSTDNNNGYNSN----SSYNSNN-- 619

Qy        705 SDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYN 764
                       || | :| ||::         :||||    |   | |||||   ||||||       |
Db        620 -------NEDDNNNNNNND-----ENCDNNNNHNNNNYNNNNNYGNNNNNNN------N 660

Qy        765 NRDGTNNSMKLYAYNSHN-------LFQPDNNKNTSNIQNINTNKNNQDGN-----VNYS 812
               |:|  ||      :  |::|             |:|  | ::   |::  |:   :| :        | |
Db        661 NKDNNNNDGNGSSNNNNNDDDDEEEEDDEDDNNNNNDDDNMSDNEEMEDNDEDNDEYNNS 720

Qy        813 MNFCHYNLNDKNYLIDLN---------NKEQKDKNI-HGCDNNIIQNRNDFEKKKKTNFY 862
               :     |  | |:  ||          |  : ||||  |      :  | || :
Db        721 NDSYKYEEKDSNHEKDLKKDIIEGDMINSVKYDKNIGHHTTNKSEISTNYFENSCNMSVN 780

Qy        863 NNNN-----------------IVIVNNNMGNNNSPRMKYGLC-------GSHTSIDNMKN 898
               |:||                 :  : |  |: ||              |              ||    :|| |
Db        781 NSNNEAYDDNCNNGFMNHDEGLTLNNGNVSNNK--------CDIIIPEDGS-VMYENMIN 831

Qy        899 NEMKNNEMKQNEMKDNHIKSNNNNSSSSSSNNNIYNNINDDDTFQNDY---------- 947
                  :  |  :   |    :|::      :||:|  |   ::::|:|||||:        :  |  |
Db        832 ---RGNGLTSNINNNNNV---SNNNSISCNADDNVYNNINN---YINTYMETTNNKNHIE 882

Qy        948 --CHNDNTFTIRRK-NNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKE---------- 994
                 |:  |:  |        :|  :||        |  |:|     |  |:    |:   |   |  ||
Db        883 NRCNQDSYSTNEEPLSNHSINDPGKIKDGIMYDGNDLD--MNGTQEHSKEEGMDVFEPNF 940

Qy        995 -----------KYTYPTLSTNEDIYNKEMEGKHIR----LDDQDK------------- 1024
                           |:  |      :    |  |     |||: |
Db        941 FELKRNSSDGQNKHLEP--GVQKKISKKRSKVKHERNSKILDDEKKEVLNKVSQITRVGG 998

Qy       1025 ------------------------------------------------------------ 1024

Db        999 VCFDKNRQRWIAHWKIDGKYHKHYFPISQYGFENARERAVSCRKQAEKLFNLPEIQPRNR 1058

Qy       1025 -------------------------YDD-------------------------------- 1027
                                       ||:
Db       1059 WNQIKVNGTSHIKKAAKLPRCEGIGYDELSQSWVSTFVVHKKFSIEELGFYEAREKAIYC 1118

Qy       1028 ---------------------------------------------------NDNNNVDNN 1036
                                                                 ::|| |:|
Db       1119 RKTFEKVNVHDDYECLLNDRLGLRNEEKDELSDLINIDKNALDNLELETSVHNNNKVKHN 1178

Qy       1037 NKNNVDNNVDNNNVDNNVDNNDKNNVDNN--NVDNDDDDVDFHNIKNFNNNEYLSYFQKN 1094
              |||  ||  |||  |||  :|||  :|:::|  :  ||        :||:::::|    ||  ::|:|:|     |:
Db       1179 NNNN--NNNNNNNNNNNNNSEKMRIKNNDFSVDNNNENVGTGEIK-ISNDKYLKITQEA 1235

Qy       1095 VDTIINNCLN-SL-DISSMYDDTKEILN-NILLSKYKAEKDNVIKKYINEDIKNMSLEEI 1151
              :: |::|  :  || :|     | ::   | | || |:          |:|   :||:|
Db       1236 IEMILSNIKHKSLPEIKMKLIDKQKFENYNTLLDKH--------FKFIT-SVKNIS---- 1282

Qy       1152 DKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDL 1192
                             | : :||  |  :              |||  |
Db       1283 -------------QLRRYISLFHKFIIYHTLPHNISLRKQL 1310
```

```
RESULT 18
US-10-422-866-4
; Sequence 4, Application US/10422866
; Publication No. US20030170857A1
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. US20030170857A1el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1054
;  CURRENT APPLICATION NUMBER: US/10/422,866
;  CURRENT FILING DATE:  2003-04-23
;  PRIOR APPLICATION NUMBER: US/09/722,129
;  PRIOR FILING DATE: 2000-11-24
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 303
;   TYPE: PRT
;   ORGANISM: Human
US-10-422-866-4

  Query Match              8.6%;  Score 589.5;  DB 12;  Length 303;
  Best Local Similarity   43.6%;  Pred. No. 1e-33;
  Matches  132;  Conservative   56;  Mismatches  108;  Indels    7;  Gaps    5;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |     || |||||  ::    :  :  | : |||:|:|::    |   |   |||
Db          4 RICVCVRKRPLNQRETTLKDLDIITVPSDNVVMVHESKQKVDLTRYLQNQTFCFDHAFDD 63

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
                  |  ||: | :||:  ::   | : :||||||||||||||  |          |   : ||:
Db         64 KASNELVYQFTAQPLVESIFRKG-MATCFAYGQTGSGKTYTMGGDFSGTAQDCSKGIYAL 122

Qy        123 AAGDIFTFL--NIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
                | |:|  |   : |:| : | :: :|:||| ||:||||   :| :    ||:| ::: |  |:
Db        123 VAQDVFLLLRNSTYEKLDLK-VYGTFFEIYGGKVYDLLNWKKKLQVLEDGNQQIQVVGLQ 181

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
                   |     ||::  :   |     |               |  ||||||:  | ||   :        || : :||
Db        182 EKEVCCVEEVLNLVEIGNSCRTSRQTPVNAHSSRSHAVFQIILK--SGRIMHGKFSLVDL 239

Qy        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
              ||:||||||    ::: | :|| ||:|||||||||  |:    :|  |  |||  |:|| |||| |:
Db        240 AGNERGADTTKASRKRQLEGAEINKSLLALKECILALGQNKPHTPFRASKLTLVLRDSFI 299

Qy        301 GKS 303
              |::
Db        300 GQN 302

RESULT 19
US-10-422-866-2
; Sequence 2, Application US/10422866
; Publication No. US20030170857A1
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. US20030170857A1el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1054
;  CURRENT APPLICATION NUMBER: US/10/422,866
;  CURRENT FILING DATE:  2003-04-23
;  PRIOR APPLICATION NUMBER: US/09/722,129
;  PRIOR FILING DATE: 2000-11-24
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 492
;   TYPE: PRT
;   ORGANISM: Human
US-10-422-866-2

  Query Match              8.6%;  Score 589.5;  DB 12;  Length 492;
  Best Local Similarity   43.6%;  Pred. No. 2e-33;
  Matches  132;  Conservative   56;  Mismatches  108;  Indels    7;  Gaps    5;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |     || |||||  ::    :  :  | : |||:|:|::    |   |   |||
Db        193 RICVCVRKRPLNQRETTLKDLDIITVPSDNVVMVHESKQKVDLTRYLQNQTFCFDHAFDD 252

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
                  |  ||: | :||:  ::   | : :||||||||||||||  |          |   : ||:
Db        253 KASNELVYQFTAQPLVESIFRKG-MATCFAYGQTGSGKTYTMGGDFSGTAQDCSKGIYAL 311

Qy        123 AAGDIFTFL--NIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
                | |:|  |   : |:| : | :: :|:||| ||:||||   :| :    ||:| ::: |  |:
Db        312 VAQDVFLLLRNSTYEKLDLK-VYGTFFEIYGGKVYDLLNWKKKLQVLEDGNQQIQVVGLQ 370

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
                   |     ||::  :   |     |               |  ||||||:  | ||   :        || : :||
Db        371 EKEVCCVEEVLNLVEIGNSCRTSRQTPVNAHSSRSHAVFQIILK--SGRIMHGKFSLVDL 428

Qy        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
              ||:||||||    ::: | :|| ||:|||||||||  |:    :|  |  |||  |:|| |||| |:
Db        429 AGNERGADTTKASRKRQLEGAEINKSLLALKECILALGQNKPHTPFRASKLTLVLRDSFI 488

Qy        301 GKS 303
              |::
Db        489 GQN 491

RESULT 20
US-10-304-095-6
; Sequence 6, Application US/10304095
; Publication No. US20030134275A1
; GENERAL INFORMATION:
;  APPLICANT: Long, David M.
;  APPLICANT:  Metz, Anneke M.
;  APPLICANT:  Love, Ruschelle A.
;  TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
;  FILE REFERENCE: 47714-5009-US
;  CURRENT APPLICATION NUMBER: US/10/304,095
;  CURRENT FILING DATE:  2002-11-26
;  PRIOR APPLICATION NUMBER: US/09/417,485
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 2184
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (330)..(335)
;   OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
;   OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
US-10-304-095-6

  Query Match              7.9%;  Score 538.5;  DB 12;  Length 2184;
  Best Local Similarity   21.4%;  Pred. No. 5.6e-29;
  Matches  330;  Conservative  234;  Mismatches  489;  Indels  487;  Gaps   73;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLY----IDEPRYKVDMTKYIERH--EFI- 56
              :||    :|   :|  :    :|:: :  |:          |      |::    |   :  ||:  :   |||
Db        429 QIKCSTKKNDISHIITSRKENHLFHVQKLENNYKHPNINKQLRKTKILKYVYNYFKEFIN 488

Qy         57 ----------------------VDKVF---------DDTVDNFTVYENTIKPLIIDLYE 84
                                     : |:|               : |   :      ||   :  |  :
Db        489 NVINTKFGKIYRKFFPRKHILNKIHKIFKIIRLQIIKKYRIINIRMNRKFIKQKVYDTFF 548

Qy         85 NGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFI 144
                          |::         ||| ::      |                  |        :      |           |:  |:   |   :
Db        549 KNYDFLSFSF------KTYKIINFMVYITKKCIPIKLLGSKHNF---KIFLKNVKKFLLF 599
```

```
Qy       145 SFYEIYC-------GKLYDLLQKRKMVAALENGKKEVVVKDL-------KILRVLTKEEL 190
Db       600 NYKESFSLNQVMKNIKVKNIFQKK---ISKYNIKNRILLKNIFDNNYENKILH-RNNKEI 655

Qy       191 ILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADT- 249
Db       656 ITNINDNIKI----YNKKNDNLNNSFKIKTTLFNKLRRK-YFNKIKKINIAIQKRHLMNR 710

Qy       250 ------------------VSQNKQTQTDGANINRSLLALKECIRAMDSDKN----HIPF 286
Db       711 LIYFLFNYFIMPLIRRFFFLTKSEQT------LHKTIFFDRKIWNHFTKISNFCLYHQIF 764

Qy       287 RDSELTK-------VLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKST 339
Db       765 RNKKLKKRNEPKMDYVQNMFNVKKKG--------------EKIKTNKYIF-IKKMKKKST 809

Qy       340 --CINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRG------- 389
Db       810 NKCINNK---FSKKCIPKKKKKNLYNITRHNNIFIKKDMEKKSKTNNLINKSIDNLYKLK 866

Qy       390 KINDKI-------------------------------ERNNILKNKSF----------DK 408
Db       867 EINKKSVRPYIKKFYYKIKKKYFALKKMYIHMRMAKEEKSNIKLERAFKHFFIFAQEKEH 926

Qy       409 PREGFTSTFGKYSSLND-------IDKIK----------KNKKK---GLINYKSTLYND 447
Db       927 ILKYFSSHFFQNRKINYGKRFNKLIHRIKNIIIKQNSGIVKNKDKTFLHLIKNKSNKNNN 986

Qy       448 NTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNN 507
Db       987 NKKKNKNNYNNNNINNNNNNNNNNNINNNNNNKCKLSN----------SKRYNIRNNNNN 1036

Qy       508 NNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSS 567
Db      1037 KKAKNNEKNN-----------------IDDSN--LEKKKKKIYIY--------KIKNIIE 1069

Qy       568 CDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPK-------D 620
Db      1070 KRNFMLK----LNSINHFISKKLRI-NWIPKKKG----LRPLINLSTLNVPEIVKQRIFE 1120

Qy       621 ILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNM 680
Db      1121 ILKSKKSSEFYFHNILNNLER------EKKDKNIKKRKKYNK--KNFNPVSLNNICNFSL 1172

Qy       681 NSINNSNMNSNSIYKSNYNSNQSISDVQI-----------------RYVNEMDTSNKNND 723
Db      1173 KCLGNMRHNNNSLFK---NTLTKTGEIELKLKKWLHYLKNWFYKKKRMKKYIKNKLKNNK 1229

Qy       724 NIF-FDAISCDNNMYPNITNNNNNNN------NNNNNNNNNIDVENYNNRDGTNNSMKLY 776
Db      1230 KIYAYICIGDFSNCYEHI-----NHNYLFKILKNFFDNINNFEFIYLFKR-----SFRLY 1279

Qy       777 AYNSHNLF---QPDNNKN--TSNIQNIN--TNKNNQDGNVNYSMN--FCHYNLNDKNYLI 827
Db      1280 NKNLNNSFLSYYPVNVKSFGLHYIRNLRELIIKSHLNDNHHFLLNQMFKTKSKSDLYIFA 1339

Qy       828 DLNNKEQKDKN---------IHGCDNNIIQNRNDFEKKKKTNFY------NNNNIVIVN- 871
Db      1340 DSYKSLQVDKRDIFMTIITVIRYYYLNIYFSIKEFKLNRKNIFYFQIFQENQMKGVYLSV 1399

Qy       872 -------------NNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIK 917
Db      1400 RDKKRVENIKKWYLNSMKKINHDEILESLKNSSININN------KNFMICTNHEQDTEEK 1453

Qy       918 SNNNNSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYT 977
Db      1454 GNTQNKEKHDIYIGPIYNNSFDSTT----TTHSSNNY---KGNNIHVSGD-YKNDGLLHK 1505

Qy       978 INSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNN 1037
Db      1506 GNN--------------------SMNE-CYVKDIK------------CNNNNNNNNNN 1531

Qy      1038 KNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDT 1097
Db      1532 NNNINNSY---NKLNCVTNNSKNDI-----------IKYHKTIDTDNSKNHTYFK----- 1572

Qy      1098 IINNCLNSLD---ISSMYDDTKEI-LNNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDK 1153
Db      1573 --NKFLNFLDKKIISNIYGLPQGFSLSNILCSLYYAYLDK------NEEFQNL------- 1617

Qy      1154 TAQSIYEKRKVLLTKLLLLFKKNVDTQINNE------TSDLRKDLVMCHICNNNP----- 1202
Db      1618 ----LYSEK-----------------QINNKYFLANGT---------CNYFNLNSLILRF 1647

Qy      1203 DDQFHFYAYSRLEKDII-NLIMLRQIWCESEN--------LRLLYQ--FLVVEYQNKSAN 1251
Db      1648 IDDFLFITLNKKNIKIFKNLLLKKKIWGSNINSSKTKIFKIPLIYKNDLLIYNFQNKYQK 1707

Qy      1252 SVLLNVSSNNGDIILLNKKLVQDNIKNSMDHN---NIHKK 1288
Db      1708 K----------KYKIKNKKKIQ-SVRNKRIHNQLVNANKK 1736
```

RESULT 21
US-10-155-533-9
; Sequence 9, Application US/10155533
; Publication No. US20030104003A1
; GENERAL INFORMATION:
; APPLICANT: Nguyen, Thanh V.
; APPLICANT: James, Anthony A.
; TITLE OF INVENTION: A No. US20030104003A1el Surface Protein of the Malaria Parasite
; FILE REFERENCE: 48417/CAB/R2682
; CURRENT APPLICATION NUMBER: US/10/155,533
; CURRENT FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1610
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-155-533-9

Query Match 7.6%; Score 518.5; DB 15; Length 1610;
Best Local Similarity 23.3%; Pred. No. 9.5e-28;
Matches 241; Conservative 164; Mismatches 348; Indels 283; Gaps 48;

```
Qy       281 KNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTC 340
Db       103 RNKLNFHNIQTDNT---IYKPKKKYYEVGKVREKIK-----MYTLFKDDKINTLK----- 149

Qy       341 INEEDDTNTERISILDS----KG-SEMNASSIENV-------VIKSNHLLSNNNNNKINR 388
Db       150 CNEEESVTSETINKHDSSVKVKGRRKKNITSNENIMNPNNKSVSSINTNLSDSSNNK-ND 208

Qy       389 GKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDN 448
Db       209 NSLNSK-KNDNTLNSKKNDN--------------------TLNSKK----------NDN 236

Qy       449 TINKKHNNNN-NNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMI---NNNINNNINVNNN 504
Db       237 TLNSKNNNNSLNSKNNNNSLNSKKSNNSYNDKH------IDHIIPEGKNKINNNIDVKHN 290

Qy       505 NNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIK--- 561
Db       291 INNKLNEINEE---------PYEDT---HNKEENSSNKNDNDEK----RKEENNNITRRY 334

Qy       562 LKNRSSCDNIMNKK-----KNNLHLARHSVGSKLTMFSYDPQK----------NKDNTFF 606
Db       335 IKNDQMSYNNINTNSNEYDKNASTLDETYIGKTFEGYVYSVNKNAACIKLKNINKYGLLF 394

Qy       607 KSNINKMED--------------------NTPKDILYESRNVSNMNGNVLLGLNKNTHH 645
Db       395 KNKANLGDDIEDMNDFFEKDQPVHVKILGINTKKNIFYLG-NIIKYNENIKLSKGEYSKG 453

Qy       646 DISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSIS 705
```

```
Db        454 LITKVCDSY-------CFIKVLKNGSTGYLHKSKLFCMNDKEKKNND--NQNYD------ 498

Qy        706 DVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNN 765

Db        499 -------------NPNNDNPNYDNQNYDNQNY------NNPNYDHPNYDNQNYDNQNYNN 539

Qy        766 RDGTNNSMKLYAYNSHNLFQPD-----------NNKNTSNIQNINTNKNNQDGNVNYSMN 814

Db        540 PNNDYSTSQL--YNSDNL-QMDFIYKLQFTKIFNIWDIIDVEILGTPQN--DYKSNYILT 594

Qy        815 FCHYNLNDK---NYL------IDLNNKEQKDKNIHGCDNN--------IIQNRNDFEKKK 857

Db        595 IPRGSKTFKKILNYLNVLKENEDINNIQYKGDYIYSIDNNKNDNIIDSDINNHHINNKKK 654

Qy        858 KTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNN----EMKDNEMKD 913

Db        655 KKNLYD------IQNNM--NHSPFNKF-----HTEDEYLFNDHVQENVHTFYEKNKKYKI 701

Qy        914 NHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDD 973

Db        702 TYDKENNH------KMNKSYYLKKNKELPFNNKF--------------KKIIKNIY---D 738

Qy        974 IIYTIN-------------SLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLD 1020

Db        739 LPNTISLSMLSKTIKIPLASIKKYF---IIHENKEYNSSYKINSEQI---KRICQHFKID 792

Qy       1021 -DQDKYDDNDNNNVDNNNKN----------------------NVDNNVDNNN--VDNNV 1054

Db        793 CNVEQRDDNVVTKVNGTTKDCQEKVFKNVTQDRLKEGEQERVIKVEAKIKNDEMVMQEQK 852

Qy       1055 DNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDD 1114

Db        853 DTKEEKHMDVQFIEEKDINVQHINVQDMDVQD-MDVQDINVQDMDVQNINVQDINIQDMD 911

Qy       1115 TKEILNNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFK 1174

Db        912 VQNINNSITLNKSTSCQTDESRDAPGGD-QNESLDEKDSMEKSKEKKKKKGKSR-----K 965

Qy       1175 KNVDTQINNETSDLRK 1190

Db        966 KNKDTNLTLKSDSIQK 981
```

RESULT 22
US-10-161-051-119
; Sequence 119, Application US/10161051
; Publication No. US20030152945A1
; GENERAL INFORMATION:
; APPLICANT: Peter Deak
; APPLICANT: David M Glover
; APPLICANT: Carol Midgley
; TITLE OF INVENTION: Cell cycle progression proteins
; FILE REFERENCE: CCI-021CP
; CURRENT APPLICATION NUMBER: US/10/161,051
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: GB 0007268.6
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 119
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-161-051-119

```
  Query Match             7.2%;  Score 491;  DB 12;  Length 464;
  Best Local Similarity   31.3%;  Pred. No. 1.6e-26;
  Matches  152;  Conservative   73;  Mismatches  169;  Indels   92;  Gaps   17;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYID------------EPRYKVDMTK 48

Db          6 HTNIKVAVRVRPYNVRELEQKQRSIIKVMDRSALLFDPDEEDDEFFFQGAKQPYR-DITK 64

Qy         49 YIERH---EFIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTM 105

Db         65 RMNKKLTMEF--DRVFDIDNSNQDLFEECTAPL-VDAVLNGYNCSVFVYGATGAGKTFTM 121

Qy        106 LGSQPYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVA 165

Db        122 LGSEAH-----PGLTYLTMQDLFDKIQAQSDVRKFDVGVSYLEVYNEHVMNLLTKSGPLK 176

Qy        166 ALE--NGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDL 223

Db        177 LREDNNG---VVVSGLCLTPIYSAEELLRMLMLGNSHRTQHPTDANAESSRSHAIFQVHI 233

Qy        224 KDINKNTSLG---KIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSD 280

Db        234 RITERKTDTKRTVKLSMIDLAGSERAASTKGIGVRFK-EGASINKSLLALGNCINKLADG 292

Qy        281 KNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTC 340

Db        293 LKHIPYRDSNLTRILKDSLGGNCRTLMVANVSMSSLTYEDTYNTLKYASRAKKIR----- 347

Qy        341 INEEDDTNTERISILDSK-GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNN 399

Db        348 -------TTLKQNVLKSKMPTEFYVKKIDEVVAENERLK-----------------ERNK 383

Qy        400 ILKNKSFDKPREGFTSTFGK------YSSLNDI---------------DKIKKNKKKGLI 438

Db        384 ALEAKATQLERAG-NSGFDPLELKTWYSKIDAVYAAARQLQEHVLGMRSKIKN------I 436

Qy        439 NYKSTL 444

Db        437 NYRQTL 442
```

RESULT 23
US-10-146-473-82
; Sequence 82, Application US/10146473
; Publication No. US20030108888A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Gout, Ivan
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Gure, Ali
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Old, Lloyd
; TITLE OF INVENTION: Breast Cancer Antigens
; FILE REFERENCE: L00461/70130(JRV)
; CURRENT APPLICATION NUMBER: US/10/146,473
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/291,150
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 82
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-146-473-82

```
  Query Match             6.7%;  Score 460.5;  DB 15;  Length 1388;
  Best Local Similarity   18.5%;  Pred. No. 9.6e-24;
  Matches  286;  Conservative  262;  Mismatches  517;  Indels  477;  Gaps   65;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDT 64

Db         27 IKVFVRIRPPAE-RSGSADGE----QNLCLSVLSSTSLRLHSNP--EPKTFTFDHVADVD 79

Qy         65 VDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDT-------- 116

Db         80 TTQESVFA-TVAKSIVESCMSGYNGTIFAYGQTGSGKTFTMMGP---SESDNFSHNLRGV 135

Qy        117 -PGIFQYAAGDIFTFLNIYDKDNTKG----IFI---SFYEIYCGKLYDLLQKRKMVAAL- 167
```

```
Db        136 IPRSFEY-------LFSLIDREKEKAGAGKSFLCKCSFIEIYNEQIYDLLDSASAGLYLR 188

Qy        168 ENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDIN 227
Db        189 EHIKKGVFVVGAVEQVVTSAAEAYQVLSGGWRNRRVASTSMNRESSRSHAVFTITIESME 248

Qy        228 KNTSLGKIA-----FIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM----D 278
Db        249 KSNEIVNIRTSLLNLVDLAGSERQKDTHAEGMRLKEAG-NINRSLSCLGQVITALVDVGN 307

Qy        279 SDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKS 338
Db        308 GKQRHVCYRDSKLTFLLRDSLGGNAKTAIIANVHPGSRCFGETLSTLNFAQRAKLIKNKA 367

Qy        339 TCINEEDDTNTERI---------------------------------------SILDSK 358
Db        368 V-VNEDTQGNVSQLQAEVKRLKEQLAELASGQTPPESFLTRDKKKTNYMEYFQEAMLFFK 426

Qy        359 GSEMNASS-------IENVVIKSNHLLSNNNN---------------NKINRG------- 389
Db        427 KSEQEKKSLIEKVTQLEDLTLKKEKFIQSNKMIVKFREDQIIRLEKLHKESRGGFLPEEQ 486

Qy        390 ------------KINDKIE----------RNNILKNKS--------FDKPREGFTSTFGK 419
Db        487 DRLLSELRNEIQTLREQIEHHPRVAKYAMENHSLREENRRLRLLEPVKRAQEMDAQTIAK 546

Qy        420 ----YSSLNDIDKIKKNK--------------------KKGLINYKSTLYNDNTINKKHN 455
Db        547 LEKAFSEISGMEKSDKNQQGFSPKAQKEPCLFANTEKLKAQLLQIQTEL----------- 595

Qy        456 NNNNNNNDNNDNNNDNNNNNNNNDSSSMVNNMINHMINNNINNNINVNN--NNNNNNNNN 513
Db        596 ---NNSKQEYEEFKELTRKRQLELESELQSL--QKANLNLENLLEATKACKRQEVSQLNK 650

Qy        514 SHNNHL---------------PQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRD 558
Db        651 IHAETLKIITTPTKAYQLHSRPVPKLS-PEMGSFGSLYTQNSSILDND--------ILNE 701

Qy        559 NIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKME-DNT 617
Db        702 PVPPEMNEQAFEAISEELRTVQEQMSALQAKL-----DEEEHK-NLKLQQHVDKLEHHST 755

Qy        618 PKDILYESRNV--SNMNGNVLLGLN--------KNTHHDISTKDENHNDNKINNGVINII 667
Db        756 QMQELFSSERIDWTKQQEELLSQLNVLEKQLQETQTKNDF-LKSEVHD--------LRVV 806

Qy        668 NNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQ-----SISDVQIRYVNEMDTSNKNN 722
Db        807 LHS-------------ADKELSSVKLEYSSFKTNQEKEFNKLSERHMHVQLQLDNLRLEN 853

Qy        723 DNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHN 782
Db        854 EKL-LESKACLQDSYDNLQEIMKFEIDQLSRNLQNFKKENETLKSDLNNLMELLEAEKER 912

Qy        783 LFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGC 842
Db        913 -----NNKLSLQFEEDKENSSKEILKVLEAV------------------RQEKQKETAKC 949

Qy        843 DNNI-------------------IQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMK 883
Db        950 EQQMAKVQKLEESLLATEKVISSLEKSRDSDKK------------VVADLM--NQIQELR 995

Qy        884 YGLCGSHTSIDNMKNNEMKNNEMKDN---------------------EMKDN-------- 914
Db        996 TSVCEKTETIDTLK-QELKDINCKYNSALVDREESRVLIKKQEVDILDLKETLRLRILSE 1054

Qy        915 ------------HIKSNNNNSSSSSSSNNNIYNNINDD----DTFQNDYCHNDNTFTIRR 958
Db       1055 DIERDMLCEDLAHATEQLNMLTEASKKHSGLLQSAQEELTKKEALIQELQHKLN----QK 1110

Qy        959 KNNTNINSNIYQNDDIIYTINSLNDYMSN----------TLLHFKE-------------- 994
Db       1111 KEEVEQKKNEYN-----FKMRQLEHVMDSAAEDPQSPKTPPHFQTHLAKLLETQEQEIED 1165

Qy        995 --------KYTYPTLSTNEDIYNKEM--------EGKHIRLDDQDKYDDNDNNNVDNNNK 1038
Db       1166 GRASKTSLEHLVTKLNEDREVKNAEILRMKEQLREMENLRLESQQLIEKN---------- 1215

Qy       1039 NNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTI 1098
Db       1216 WLLQGQLD--DIKRQKENSDQNHPDNQQLKNEQEE----SIK-----ERLA--------- 1255

Qy       1099 INNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSL-EEIDKT--- 1154
Db       1256 ----------------KSKIVEEML--KMKADLEEVQSALYNKEMECLRMTDEVERTQTL 1297

Qy       1155 AQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMC-------HICNNNPDDQFH 1207
Db       1298 ESKAFQEKEQLRSKLEEMYEERERT--SQEMEMLRKQ-VECLAEENGKLVGHQNLHQKIQ 1354

Qy       1208 FYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKS 1249
Db       1355 YVV--RLKKENVRLAE------ETEKLRAENVFLKEKKRSES 1388
```

RESULT 24
US-10-342-224-62
; Sequence 62, Application US/10342224
; Publication No. US20030162294A1
; GENERAL INFORMATION:
; APPLICANT: Nathalie Verbruggen
; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress
; FILE REFERENCE: CNN-012US
; CURRENT APPLICATION NUMBER: US/10/342,224
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US/09/762,154
; PRIOR FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: EP 98202634.6
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 1269
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-342-224-62

Query Match 6.6%; Score 454.5; DB 12; Length 1269;
Best Local Similarity 33.3%; Pred. No. 2.3e-23;
Matches 129; Conservative 80; Mismatches 137; Indels 41; Gaps 15;

```
Qy          1 MNSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKV 60
Db        894 MKGKIRVYCRIRPLNEKESSEREKQMLTTVDEFTV---EHAWKDD-----KRKQHIYDRV 945

Qy         61 FDDTVDNFTVYENTIKPLI---IDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTP 117
Db        946 FDMRASQDDIFEDT-KYLVQSAVDGY-NVCI---FAYGQTGSGKTFTI-----YGHESNP 995

Qy        118 GIFQYAAGDIFTFLNIYDKDNTKGIF-ISFY--EIYCGKLYDLL----QKRKMVAALENG 170
Db        996 GLTPRATKELF---NILKRDSKRFSFSLKAYMVELYQDTLVDLLLPKSARRLKLEIKKDS 1052

Qy        171 KKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNT 230
Db       1053 KGMVFVENVTTIPISTLEELRMILERGSERRHVSGTNMNEESSRSHLILSVVIESIDLQT 1112

Qy        231 ---SLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFR 287
Db       1113 QSAARGKLSFVDLAGSERVKKSGSAGCQLK-EAQSINKSLSALGDVIGALSSGNQHIPYR 1171

Qy        288 DSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKN-KSTCINEEDD 346
Db       1172 NHKLTMLMSDSLGGNAKTLMFVNVSPAESNLDETYNSLLYASRVRTIVNDPSKHISSKEM 1231

Qy        347 TNTERI-----SILDSKGSEMNASSIE 368
```

```
                :::              || | :   ||
Db       1232 VRLKKLVAYWKEQAGKKGEEEDLVDIE 1258

RESULT 25
US-09-883-096-5
; Sequence 5, Application US/09883096
; Patent No. US20020110883A1
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Craven, Andrew
;  APPLICANT:  Yu, Ming
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Patel, Umesh A.
;  APPLICANT:  Davies, Katherine A.
;  TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR USE
;  FILE REFERENCE: 020552-001410US
;  CURRENT APPLICATION NUMBER: US/09/883,096
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: US 09/594,655
;  PRIOR FILING DATE: 2000-06-15
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 383
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: HsKip3a
;   OTHER INFORMATION: fragment
;   OTHER INFORMATION: Amino acid sequence of HsKip3a fragment used in
;   OTHER INFORMATION: the ATPase assay (Figure 4).
US-09-883-096-5

  Query Match              6.6%;  Score 453;  DB 10;  Length 383;
  Best Local Similarity   32.8%;  Pred. No. 5.9e-24;
  Matches  117;  Conservative   72;  Mismatches  140;  Indels   28;  Gaps    7;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYI--DEPRYKVDMTKYIERHE----- 54
              :| ::|||| || :   |    :    :: | :    |      :||        |:   |:
Db          5 DSTLQVVVRVRPPTPRELDSQRRPVVQVVDERVLVFNPEEPDGGFPGLKWGGTHDGPKKK 64

Qy         55 -----FIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQ 109
                   |: |:|| :       |:::|      ::| :   |   || ||||  ||:|||:|||
Db         65 GKDLTFVFDRVFGEAATQQDVFQHTTHS-VLDSFLQGYNCSVFAYGATGAGKTHTML--- 120

Qy        110 PYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALEN 169
                |:   |||         :::    |       ::        : ||: |:|   :::|||: :   :|   |:
Db        121 --GREGDPGIMYLTTVELYRRLEARQQEKHFEVLISYQEVYNEQIHDLLEPKGPLAIRED 178

Qy        170 GKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKD---- 225
                 | |||:   |     :  :  |:|:  :     |         |        ||||||    | :|
Db        179 PDKGVVVQGLSFHQPASAEQLLEILTRGNRNRTQHPTDANATSSRSHAIFQIFVKQQDRV 238

Qy        226 --INKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM---DSD 280
                : :    : |:: |||||||| :  | :: :: : :||||||||||       :   |:
Db        239 PGLTQAVQVAKMSLIDLAGSERASSTHAKGERLR-EGANINRSLLALINVLNALADAKGR 297

Qy        281 KNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNK 337
              | |:|:|||:||::|:|    |    :::|||  |||:         |  |  |||:|:  |  |     :  |
Db        298 KTHVPYRDSKLTRLLKDSLGGNCRTVMIAAISPSSLTYEDTYNTLKYADRAKEIRLK 354

RESULT 26
US-09-883-096-2
; Sequence 2, Application US/09883096
; Patent No. US20020110883A1
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Craven, Andrew
;  APPLICANT:  Yu, Ming
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Patel, Umesh A.
;  APPLICANT:  Davies, Katherine A.
;  TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR USE
;  FILE REFERENCE: 020552-001410US
;  CURRENT APPLICATION NUMBER: US/09/883,096
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: US 09/594,655
;  PRIOR FILING DATE: 2000-06-15
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 864
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Amino acid sequence encoded by human kinesin motor
;   OTHER INFORMATION: protein gene HsKip3a (Figure 1).
;   OTHER INFORMATION: Description of Artificial Sequence:Amino acid
;   OTHER INFORMATION: sequence of HsKip3a.
US-09-883-096-2

  Query Match              6.6%;  Score 451;  DB 10;  Length 864;
  Best Local Similarity   32.7%;  Pred. No. 2.4e-23;
  Matches  116;  Conservative   72;  Mismatches  139;  Indels   28;  Gaps    7;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYI--DEPRYKVDMTKYIERHE----- 54
              :| ::|||| || :   |    :    :: | :    |      :||        |:   |:
Db          5 DSTLQVVVRVRPPTPRELDSQRRPVVQVVDERVLVFNPEEPDGGFPGLKWGGTHDGPKKK 64

Qy         55 -----FIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQ 109
                   |: |:|| :       |:::|      ::| :   |   || ||||  ||:|||:|||
Db         65 GKDLTFVFDRVFGEAATQQDVFQHTTHS-VLDSFLQGYNCSVFAYGATGAGKTHTML--- 120

Qy        110 PYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALEN 169
                |:   |||         :::    |       ::        : ||: |:|   :::|||: :   :|   |:
Db        121 --GREGDPGIMYLTTVELYRRLEARQQEKHFEVLISYQEVYNEQIHDLLEPKGPLAIRED 178

Qy        170 GKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKD---- 225
                 | |||:   |     :  :  |:|:  :     |         |        ||||||    | :|
Db        179 PDKGVVVQGLSFHQPASAEQLLEILTRGNRNRTQHPTDANATSSRSHAIFQIFVKQQDRV 238

Qy        226 --INKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM---DSD 280
                : :    : |:: |||||||| :  | :: :: : :||||||||||       :   |:
Db        239 PGLTQAVQVAKMSLIDLAGSERASSTHAKGERLR-EGANINRSLLALINVLNALADAKGR 297

Qy        281 KNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFK 335
              | |:|:|||:||::|:|    |    :::|||  |||:         |  |  |||:|:  |  |     :
Db        298 KTHVPYRDSKLTRLLKDSLGGNCRTVMIAAISPSSLTYEDTYNTLKYADRAKEIR 352

RESULT 27
US-10-157-031-38
; Sequence 38, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
;  APPLICANT: Baranova, A. V.
;  APPLICANT:  Yankovsky, N. K.
;  APPLICANT:  Kozlov, A. P.
;  APPLICANT:  Lobashev, A. V.
;  APPLICANT:  Krukovskaya, L. L.
;  TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
;  FILE REFERENCE: 2760-103
;  CURRENT APPLICATION NUMBER: US/10/157,031
;  CURRENT FILING DATE:  2002-05-30
;  NUMBER OF SEQ ID NOS: 415
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
;   LENGTH: 1648
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-157-031-38

  Query Match              6.5%;  Score 446;  DB 15;  Length 1648;
  Best Local Similarity   19.8%;  Pred. No. 1.3e-22;
```

```
  Matches  284;  Conservative  213;  Mismatches  507;  Indels  428;  Gaps   59;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDI-------ITVKNNCTLYIDEPRYKVDMTKYIERHE 54
Db        356 NSQVTVAVRVRPFTKREKIEKASQVVFMSGKEITVEHPDTKQVYNFIYDVSFWSFDECHP 415

Qy         55 FIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQS 114
Db        416 HYASQT--------TVYEKLAAPLLERAFEGFNTC-LFAYGQTGSGKSYTMMGF-----S 461

Qy        115 DTPGIFQYAAGDIFT-FLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKE 173
Db        462 EEPGIIPRFCEDLFSQVARKQTQEVSYHIEMSFFEVYNEKIHDLL-----VCKDENGQRK 516

Qy        174 VVVKDLKILRVLTK-------EELILKMID----------GVLLRKIGVNSQNDESSRS 215
Db        517 ------QPLRVREHPVYGPYVEALSMNIVSSYADIQSWLELGNKQRATAATGMNDKSSRS 570

Qy        216 HAILNIDLKDINKNTSLG---------KIAFIDLAGSERGADTVSQNKQTQTDGANINRS 266
Db        571 HSVFTLVMTQTKTEFVEGEEHDHRITSRINLIDLAGSER-CSTAHTNGDRLKEGVSINKS 629

Qy        267 LLALKECIRAMDSDKNH----IPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTL 322
Db        630 LLTLGKVISALSEQANQRSVFIPYRESVLTWLLKESLGGNSKTAMIATISPAASNIEETL 689

Qy        323 NTLRYSSRVKNFKNKSTCINEEDDTNTERI-----SILDSKGSEMNASSIENVVIKSNHL 377
Db        690 STLRYANQARLIVNIAK-VNE--DMNAKLIRELKAEIAKLKAAQRNSRNID----PERYR 742

Qy        378 LSNNNNNKINRGKIN----DKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNK 433
Db        743 LCRQEITSL-RMKLHQQERDMAEMQRVWKEK-FEQAEK------------RKLQETKELQ 788

Qy        434 KKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINN 493
Db        789 KAGIM-----------------------FQMDNHLPNLVNLNED--PQLSEMLLYMIKE 822

Qy        494 NINNNINVNNNNNNNNNNNSH----------NNHLPQPNYAFTDTSDFSSLDDMNCHLN 543
Db        823 --------GTTTVGKYKPNSSHDIQLSGVLIADDHCTIKN--FGGTVSIIPVGEAKTYVN 872

Qy        544 -----------NNDKSI--------FLHKKNLRDNIKLKNRSSCDNIMNK-------KKN 577
Db        873 GKHILEITVLRHGDRVILGGDHYFRFNHPVEVQ---KGKRPSGRDTPISEGPKDFEFAKN 929

Qy        578 NLHLARHSVGSKLTMFSYDPQ-KNKDNTFFKSNINK---MEDNTPKDILYESRNVSNMNG 633
Db        930 ELLMAQR---SQLEAEIKEAQLKAKEEMMQGIQIAKEMAQQELSSQKAAYESK------- 979

Qy        634 NVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSI 693
Db        980 ------IKALEAELREESQRKKMQEINNQKA----NHKIEELEKAKQHLEQEIYVNKKRL 1029

Qy        694 YKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNN 753
Db       1030 EMETLATKQALEDHSIRHARILEALETEKQKIAKEV------------------------ 1065

Qy        754 NNNNIDVENYNNRDGTN------NSMKL-----------------YAYNSHNLFQPDNNK 790
Db       1066 ---QILQQNRNNRDKTFTVQTTWSSMKLSMMIQEANAISSKLKTYYVFGRHDISDKSSSD 1122

Qy        791 NTSNIQNI-----------------------NTNKNNQDGNVNYSMNFCHYNLNDKNYL 826
Db       1123 TSIRVRNLKLGISTFWSLEKFESKLAAMKELYESNGSNRGEDA----FCDPEDEWEPDI 1177

Qy        827 IDL---------NNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNN 877
Db       1178 TDAPVSSLSRRRSRSLMKNRRISGCLHDI-----QVHPIKNLHSSHSSGLMDKSSTIYSN 1232

Qy        878 NSPRMKYGLC----GS-----------HTSIDNMKNNEM----------KNNEMKDNEM 911
Db       1233 SAESFLPGICKELIGSSLDFFGQSYDEERTIADSLINSFLKIYNGLFAISKAHEEQDEES 1292

Qy        912 KDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNI--NSNIY 969
Db       1293 QDNLFSSDRAIQSL----------------TIQT-ACAFEQLVVLMKHWLSDLLPCTNIA 1335

Qy        970 Q-NDDIIYTINSLNDY---------------------------------------MSNTL 989
Db       1336 RLEDELRQEVKKLGGYLQLFLQGCCLDISSMIKEAQKNAIQIVQQAVKYVGQLAVLKGSK 1395

Qy        990 LHFKEKYTYPTLSTNEDIYNKEME----GKHIRLDD-QDKYDDNDNNNVDNNNKNNVDNN 1044
Db       1396 LHFLENGNNKAASVQEEFMDAVCDGVGLGMKILLDSGLEKAKELQHELFRQCTKNEVTKE 1455

Qy       1045 VDNNNVD-----NNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNN--EYLSYFQKNVDT 1097
Db       1456 MKTNAMGLIRSLENIFAESKIKSFRRQV--QEENFEYQDFKRMVNRAPEFLK-------- 1505

Qy       1098 IINNCLNSLDISSMYDDTKEILNNILLSKYK-AEKDNVIKKYINEDIKNMSLE-----EI 1151
Db       1506 -LKHCL-------------EKAIEIIISALKGCHSDINLLQTCVESIRNLASDFYSDFSV 1551

Qy       1152 DKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPD 1203
Db       1552 PSTSVGSYESR--------------VTHIVHQELESLAKSLLFCFESEESPD 1589

RESULT 28
US-09-883-096-4
; Sequence 4, Application US/09883096
; Patent No. US20020110883A1
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Craven, Andrew
;  APPLICANT:  Yu, Ming
;  APPLICANT:  Sakowicz, Roman
;  APPLICANT:  Patel, Umesh A.
;  APPLICANT:  Davies, Katherine A.
;  TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR USE
;  FILE REFERENCE: 020552-001410US
;  CURRENT APPLICATION NUMBER: US/09/883,096
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: US 09/594,655
;  PRIOR FILING DATE: 2000-06-15
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 338
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: HsKip3a fragment.
;   OTHER INFORMATION: Amino acid sequence of the motor domain fragment
;   OTHER INFORMATION: of HsKip3a (Figure 3).
US-09-883-096-4

  Query Match             6.3%;  Score 429;  DB 10;  Length 338;
  Best Local Similarity   32.8%;  Pred. No. 2.4e-22;
  Matches  113;  Conservative   69;  Mismatches  135;  Indels   28;  Gaps    7;

Qy          2 NSKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYI--DEPRYKVDMTKYIERHE----- 54
Db          1 DSTLQVVVRVRPPTPRELDSQRRPVVQVVDERVLVFNPEEPDGGFPGLKWGGTHDGPKKK 60

Qy         55 -----FIVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQ 109
Db         61 GKDLTFVFDRVFGEAATQQDVFQHTTHS-VLDSFLQGYNCSVFAYGATGAGKTHTML--- 116

Qy        110 PYGQSDTPGIFQYAAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALEN 169
Db        117 --GREGDPGIMYLTTVELYRRLEARQQEKHFEVLISYQEVYNEQIHDLLEPKGPLAIRED 174

Qy        170 GKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKD---- 225
```

```
Db        175 PDKGVVVQGLSFHQPASAEQLLEILTRGNRNRTQHPTDANATSSRSHAIFQIFVKQQDRV 234

Qy        226 --INKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM---DSD 280
                : :    : |:: ||||||| : | :: :: : :||||||||||     : |:
Db        235 PGLTQAVQVAKMSLIDLAGSERASSTHAKGERLR-EGANINRSLLALINVLNALADAKGR 293

Qy        281 KNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTL 325
              | |:|:|||:||::|:|    |   :::||| |||:     | | |||
Db        294 KTHVPYRDSKLTRLLKDSLGGNCRTVMIAAISPSSLTYEDTYNTL 338
```

RESULT 29
US-09-847-874A-1
; Sequence 1, Application US/09847874A
; Patent No. US20020127668A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN
; FILE REFERENCE: PF-0593 US
; CURRENT APPLICATION NUMBER: US/09/847,874A
; CURRENT FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 09/467,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 1103
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1281811
US-09-847-874A-1

```
  Query Match            5.8%;  Score 395.5;  DB 10;  Length 1103;
  Best Local Similarity  31.3%;  Pred. No. 2.7e-19;
  Matches  126;  Conservative   71;  Mismatches  156;  Indels   49;  Gaps   16;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIE------RHEFI 56
              :||  || ||  :   |   :   :::::  | |  |: |:   |  |       |
Db          4 ASVKVAVRVRPFNARETSQDAKCVVSMQGNTTSIIN-PKQSKDAPKSFTFDYSYWSHTST 62

Qy         57 VDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDT 116
               |  |        || :  : ::: :|    ||  ||||||:||:|||:| |  ||
Db         63 EDPQF---ASQQQVYRDIGEEMLLHAFEGYNVC-IFAYGQTGAGKSYTMMGRQEPGQQ-- 116

Qy        117 PGIFQYAAGDIFTFLNIYDKDNTK---GIFISFYEIYCGKLYDLL--QKRKMVAALENGK 171
               ||      |:|:   :  :  :  :     : :|: |||| :: |||   :  |   |:
Db        117 -GIVPQLCEDLFS--RVSENQSAQLSYSVEVSYMEIYCERVRDLLNPKSRGSLRVREHPI 173

Qy        172 KEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNI--------DL 223
                  |:||  | |  :  ::   |    |     |     :  |:  ||||||:  |        |
Db        174 LGPYVQDLSKLAVTSYADIADLMDCGNKARTVAATNMNETSSRSHAVFTIVFTQRCHDQL 233

Qy        224 KDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRA---MDSD 280
               :: :  :  ||: :||||||| ||:               :||||:||  :  |  |     ||
Db        234 TGLD-SEKVSKISLVDLAGSER-ADSSGARGMGLKEGANINKSLTTLGKVISALADMQSK 291

Qy        281 K---NHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNK 337
              |     : ||:||| || :|::   | |:: ||| :||      |:||:||||: | |   :
Db        292 KRKSDFIPYRDSVLTWLLKENLGGNSRTAMIAALSPADINYEETLSTLRYADRTKQIR-C 350

Qy        338 STCINEEDDTNTERISILDSK---------GSEMNASSIENV 370
               :  |||  | |   |  |  :              ::||::| :
Db        351 NAIINE--DPNARLIRELQEEVARLRELLMAQGLSASALEGL 390
```

RESULT 30
US-10-167-831-2
; Sequence 2, Application US/10167831
; Publication No. US20030166209A1
; GENERAL INFORMATION:
; APPLICANT: Shuman, Stewart
; APPLICANT: Ho, C. Kiong
; TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
; TITLE OF INVENTION: for Treatment of Parasitic Infections
; FILE REFERENCE: D6388D
; CURRENT APPLICATION NUMBER: US/10/167,831
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 09/752,165
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 104
; SEQ ID NO 2
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: PEPTIDE
; OTHER INFORMATION: Plasmodium falciparum RNA triphosphatase Prt1
US-10-167-831-2

```
  Query Match            5.7%;  Score 388.5;  DB 12;  Length 596;
  Best Local Similarity  21.5%;  Pred. No. 3.7e-19;
  Matches  171;  Conservative  107;  Mismatches  245;  Indels  271;  Gaps   33;

Qy        372 IKSNHLLSNNNNNKINRGKINDKIERN---NILKNKSFDK-PREGFTSTFGKYSSLNDID 427
              :  | :|: :|:  ||    |  :||         | |||:  | |              ||
Db         24 LSQNIILAFDNHENINNKDIQIEIEGRVGLVIDKNKNRIKLPINTDAIIENNYSDFQ--A 81

Qy        428 KIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMI 487
                | :    : |::|    :         :: ::|:|||||  |||:|| ||||||
Db         82 GIDRESFEYLLDYFHNMTLKKRLSIRNNDNNNNNMDNNNNNMDNNNNNN----------- 130

Qy        488 NHMINNNINNNINVNNNNNNNNNNNNNSHNN---HLPQPNYAFTDTSDFSSLDDMNCHLNN 544
                          ||:||:::|: || |    ::     |      :  |  ||::
Db        131 --------NNNIHIHNSGNNTNQTHSYDKNADDNKPTCNYSY----------------- 164

Qy        545 NDKSIFLHKKN--LRDNIKLKNRSSCDNIMNKKKNNLHLHARHSVGSKLTMFSYDPQKNKD 602
                      |||  :  | :  ::||      |    |          |  ||     ||::    :|
Db        165 -------DKKNACIYDFLELKTTKSIDKYYVIKNNN---------SRIRTTTY------- 201

Qy        603 NTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNG 662

Db        202 ------------------------------------------------------------ 201

Qy        663 VINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNN 722
                   :|:  |      :  : |:   |:|  ::|                          ||
Db        202 ----LNDDNKQETESMMIQSLQKDNLNIWNVYTG-----------------------NN 233

Qy        723 DNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHN 782
               : |  |     |:: |     |||||||| :          ||         ||  |    |    : || |
Db        234 YDYFDDDEEDDDDDY----NNNNNNNNGDTGTKTNIAT---NNTHGLTTSKSQHIYN--N 284

Qy        783 LFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNF-CHYNLNDK-----NYL---IDLNNKE 833
               |   | |  :           :||         :    | |     |  |  |   ::     ||     :|:      :
Db        285 LV--DKNDSIDYRISINIEYTKPISKLYLSKNTPVHERLKERTTFINTYLGLQVDMTKIK 342

Qy        834 QKDKNIHGCD-----NNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCG 888
              |:  ::  :              |  :     ::    || |:|: :        :  :|: ||      |:|
Db        343 TKNNELYEVEIEIPSKTIFKAMSNLRNKKDSNYLH-----FICSNLVNNIR-----GIC- 391

Qy        889 SHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQN--- 945
                    :     |:     |   |   |      |    |||  |:: :| ::   |  :|||   :
Db        392 ------SQLNVFKKSKHMLKNTM----ITKLNNN--SNNQNNLSLLPNHPNDDTISSKEK 439

Qy        946 ----DYCHN------DNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEK 995
                   ||:         |  : :  ||  :|     |    | :|       :| : :|    :    |
Db        440 EKFKKYIHSVLPIVGDYMYRVVTKNEKHIKRKI--KDQLITNKEKINIFKNNVDIRRHNK 497

Qy        996 YTYPTLSTNE-DIYNKEMEGKH------IRLDDQDKYDDND-----NNNVDNNNKNNVDN 1043
              .   : |:   ||  : ||        |          :    | ::|: |:         ||     :  || |
Db        498 KSLQTI--NEVHVENKWKAFKRGTKIEVLLCSDDEEYEQNEDVQDINNEYYDQYKNEEDT 555

Qy       1044 NVDNNNV--DNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINN 1101
```

```
                 ::  ||:   | ::||: || ||:| : :        |:||: :
Db        556 SLYINNIYMHNQINNNNNNNNDNDNKNEE-------NLKNYKD---------------- 591

Qy       1102 CLNSLDISSMYDDT 1115
                        ||||
Db        592 ---------FYDDT 596

RESULT 31
US-10-363-798-2
; Sequence 2, Application US/10363798
; Publication No. US20030180280A1
; GENERAL INFORMATION:
;  APPLICANT: Kong, Xiangyin
;  APPLICANT:  Xiao, Shangxi
;  APPLICANT:  Zhao, Guoping
;  APPLICANT:  Yu, Chuan
;  APPLICANT:  Hu, Landian
;  TITLE OF INVENTION: METHOD OF DIAGNOSING AND TREATING DENTINOGENESIS IMPERFECTA
;  TITLE OF INVENTION:  TYPE II USING DENTIN SIALOPHOSPHOPROTEIN GENE AND CODED
;  TITLE OF INVENTION:  PRODUCT THEREOF
;  FILE REFERENCE: 9548.78USWO
;  CURRENT APPLICATION NUMBER: US/10/363,798
;  CURRENT FILING DATE:  2003-03-05
;  PRIOR APPLICATION NUMBER: CN 00125042.6
;  PRIOR FILING DATE: 2000-09-05
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 1253
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-363-798-2

  Query Match             5.5%;  Score 376.5;  DB 12;  Length 1253;
  Best Local Similarity   19.2%;  Pred. No. 6.9e-18;
  Matches  159;  Conservative  193;  Mismatches  334;  Indels  141;  Gaps   19;

Qy        320 QTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLS 379
              | ||  | | ||:|   |           :: |  :      || |:   |           |
Db        345 QKLNH-RESKRVENRITK--------ESETHAVGKSQDKGIEIKGPS-----------S 383

Qy        380 NNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDI-------DKIKKN 432
                | |     || |:  |           || : |     |    :  ::        | :
Db        384 GNRNITKEVGKGNEGKE----------DKGQHGMILGKGNVKTQGEVVNIEGPGQKSEPG 433

Qy        433 KKKGLIN---------YKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNN-------- 475
               :| |               |  |  ::| ::       |::: :| |:| |::::||:
Db        434 NKVGHSNTGSDSNSDGYDSYDFDDKSMQGDDPNSSDESNGNDDANSESDNNSSSRGDASY 493

Qy        476 NNDSSSMVNNMINH--MINNNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFS 533
              |:| |      |  :      ::: :: :: : ||:::| | || : :|            :|:||
Db        494 NSDESKDNGNGSDSKGAEDDDSDSTSDTNNSDSNGNGNNGNDDNDKSDSGKGKSDSSDSD 553

Qy        534 SLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMF 593
              | |      |::| |           :     |     :  ||  :   :  ::   : :|   |   :
Db        554 SSDSS----NSSDSSDSSDSDSSDSNSSSDSDSSDSDSSDSSDSDSSDSSNSSDSSDSSD 609

Qy        594 SYDPQKNKDNTFFKSNINKME-DNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDE 652
              | |   :  |::  ||: :| | |::  |   :|                       |
Db        610 SSDSSDSSDSSDSKSDSSKSESDSSDSDSKSDS---------------------SDS 645

Qy        653 NHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIRYV 712
              | :|:  |:   :  |:||  :  ::|: :|:  ::|: :|:   |: :|:  ||
Db        646 NSSDSSDNSDSSDSSNSSNSSDSSDSSDSSDSSSSSDSSSSSDSSNSSDSSDSS------ 699

Qy        713 NEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNS 772
                   |:||  :   :              :::::::::::::|:|:  |  ::  |:  |   :::|
Db        700 ---DSSNSSESSDSSD-----------SSDSDSSDSSDSSNSNSSDSDSSNSSDSSDSS 744

Qy        773 MKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNK 832
                 : ||  :    :: |:|:  : : : :: |  :            | :| |      | :|
Db        745 DSSDSSNSSDSSDSSDSSNSSDSSDSSDSSDSS----------NSSDSN---DSSNS 791

Qy        833 EQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTS 892
                  : :   |::   : :|          ::    |:::    :::   ::||              | :|
Db        792 SDSSDSSNSSDSSNSSDSSDSSDSSDSSNSSDSSNSSDSSDSSNSSDSSDSSDSSDSS 851

Qy        893 IDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSNNNIYNNINDDDTFQNDYCHNDN 952
                  :   |     :|     ::   |:    |::::|| ||  || :   ::  :  |  :   :|     :|:
Db        852 DSDSSNRSDSSNSSDSSDSSDSSNSSDSSDSSDSSDSNESSNSSDSSDSSNSSDSDSSDS 911

Qy        953 TFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEM 1012
                       :|:: :|:    :  |    :  ||  ::    |:    :     :                |  :  |   |
Db        912 -------SNSSDSSDSSNSSDSSESSNSSDNSNSSDSSNSSDSSDSSDSSNSSDSSNSG- 963

Qy       1013 EGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDD 1072
                       |   :    | :|:|: |::: :|   :: |:::    ::  |::|  :|    ::: :|
Db        964 -------DSSNSSDSSDSNSSDSSDSSNSSDSSDSSDSSDSSDSSDSSNSSDSSDSSDSS 1016

Qy       1073 D----VDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDT 1115
              |      |   |  : :::    |       :  |:   :|   :| |  |     | :
Db       1017 DSSNSSDSSNSSDSSDSSDSSDSSDS--SNSSDSSDSSDSSDSS 1061

RESULT 32
US-10-304-095-8
; Sequence 8, Application US/10304095
; Publication No. US20030134275A1
; GENERAL INFORMATION:
;  APPLICANT: Long, David M.
;  APPLICANT:  Metz, Anneke M.
;  APPLICANT:  Love, Ruschelle A.
;  TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
;  FILE REFERENCE: 47714-5009-US
;  CURRENT APPLICATION NUMBER: US/10/304,095
;  CURRENT FILING DATE:  2002-11-26
;  PRIOR APPLICATION NUMBER: US/09/417,485
;  PRIOR FILING DATE: 2002-06-14
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 794
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-10-304-095-8

  Query Match             5.5%;  Score 374.5;  DB 12;  Length 794;
  Best Local Similarity   21.6%;  Pred. No. 5.2e-18;
  Matches  214;  Conservative  152;  Mismatches  312;  Indels  315;  Gaps   46;

Qy        326 RYSSRVKNFK----NKSTCINEEDDTNTERISILDS-KGSEMNASSIENVVIKSNHLLSN 380
                |    ||:| |      |      || ::             ||:| | |  :| :: :| :| :||
Db          9 RDKKRVENIKKWYLNSMKKINHDE--------ILESLKNSSININN-KNFMICTNH---- 55

Qy        381 NNNNKINRGKINDKIERNNI----LKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKG 436
                    :     :|    :| |:::|      : | |||           ||    :||              | |
Db         56 -EQDTEEKGNTQNK-EKHDIYIGPIYNNSFD-------STTTTHSS--------NNYKGN 98

Qy        437 LINYKSTLYNDNTINKKHNNNNN--------NNNDNNDNNNDNNNNNNNNDSSSMVNNMIN 488
                |:          ||   ::| :|:  |          |||:|:|||:|||  ||:          :|
Db         99 NIHVSGDYKNDGLLHKGNNSMNECYVKDIKCNNNNNNNNNNNNNINNS------YNKLN 152

Qy        489 HMINNNINNNINVNNNNNNNNNNNNSH-------------------------NNHLPQP 522
                  : ||:  |: |    :    : :|: |:::                              :| |
Db        153 CVTNNSKNDIIKYHKTIDTDNSKNHTYFKNKFLNFLDKKIISNIYGLPQGFSLSNILCSL 212

Qy        523 NYAFTDTSDFSS--------------LDDMNCHLNNNDKSIF----------LHKKNLR- 557
                 ||: |  :: |                   | :   |:   |  :   |               |:|||::
Db        213 YYAYLDKNEESQNLLYSEKQINNKYFLANGTCNYFNLNSLILRFIDDFLFITLNKKNIKI 272

Qy        558 -DNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDN 616
                |:  ||  :       ||  |   |     :                    :|  |:::   :     :       |  |::
Db        273 FKNLLLKKKIWGSNI-NSSKTKIFKIPLIYKNDLLIYNFQNKYQQKKKYKIKNKKKIQSV 331

Qy        617 TPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSIN 676
```

```
Db        332 RNKRI-----------HNQLVNANKKKHTSVQ-KD------KINK-YINLIHPTIQK--N 370

Qy        677 NSNMNSINNSNMNSNSIY-KSNYNSNQSI-SDVQIRYVNEMDTSNKNNDNIFFDAISCDN 734

Db        371 DSVLSS--NSIMNFERIYNKESHKSNSSIRTDIPNSVVNDDIEYNQKSDNNSYST----N 424

Qy        735 NMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLF-QPDNNKNTS 793

Db        425 NLY-------NNINMTQNGDNNNVNIFKHVQND------SFQCFNSNNLYIEKDIKEN-- 469

Qy        794 NIQNIN---------TNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDN 844

Db        470 NISQINRKLCTKRNFTKKSRKINTVTY--------LQIDKVIKILKCKKKYIKHIK--KM 519

Qy        845 NIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNN 904

Db        520 KYMNNFQNFKKLKKLQKFQNASF-------------------------------ELKIN 547

Qy        905 EMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNI 964

Db        548 KINKNIRRLNKLKKRKNHSI----------------------------------NITPV 572

Qy        965 NSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDK 1024

Db        573 TSIEWLNNS--YTFDFINNSIQSTSYPWKNK-------CDATIRN------HLHL----- 612

Qy       1025 YDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDVDFHNIKNFNN 1084

Db        613 ----HNVIIDKNNKTYFMKNLVENRIVRNIISKQKK---------------CQSLYKNKQ 653

Qy       1085 NEYLSYFQKNVDTIINN-------CLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKK 1137

Db        654 NVYFCY--KNNFSLLKSSILKFICCIKTL---------KKMFNAFTNSTYNTKFILFLIS 702

Qy       1138 YINED-IKNMSLEEID----KTAQSIYEKRKVL-----LTKLLLLFKKNVDTQINNETSD 1187

Db        703 YMNKMLIKNKKLKFVKLFLIQTAIEAFRYARIFNQQDSFYPCLQHFRKIKKRLINKYKIG 762

Qy       1188 LRKDLVMCHICNNNPDDQFHFYAYSRLEKDIIN 1220

Db        763 HNKNLL-----------REFFFLFNFIKKELYN 784
```

RESULT 33
US-10-087-464-47
; Sequence 47, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 47
; LENGTH: 1191
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-47

```
  Query Match             5.5%;  Score 373;  DB 15;  Length 1191;
  Best Local Similarity   21.0%;  Pred. No. 1.1e-17;
  Matches  169;  Conservative  137;  Mismatches  371;  Indels  128;  Gaps   29;

Qy        324 TLRYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENV---VIKSNHLLSN 380

Db         72 SLGYELSEKDEGNKNDLINSATDVSTE----------------LENLKERLFPELELYTN 115

Qy        381 NNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLIN- 439

Db        116 DQNSRNNTPNL-----RKGSLGFDSFKKLELGTLNQFDKDKMINLKDETNMNEFEGFLGR 170

Qy        440 -------YKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMIN 492

Db        171 NSMASNVVTSELFDEPVDDSSSTTTSTGTKLQNVPSNDNNGELLKDEP----------ID 220

Qy        493 NNINNNINVNNNNN---NNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNN---D 546

Db        221 DYINNNSKVESEDNYYAQQNMQSQSKDNYASEQNVADQSTDNYPTQHDVPVQLRDNYASE 280

Qy        547 KSIFLHKKNLRD-------NIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYD--- 596

Db        281 QEYFDRGEQLNDVSADNNTSNKLKDEPVDNNTSNKLKD--EPVDNNTSNKLKDEPVDDNT 338

Qy        597 PQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDE---N 653

Db        339 SNKLKDEPVDNNTINKLKDEP-----------VDDNTSNILKDEPVDDHAGKHLKDEPVDD 388

Qy        654 HNDNKINNGVINIINNSNVNSINNSNMNSINNS-----NMNSNSIYKSNYNSNQSISDVQ 708

Db        389 HAGKHMKDEPVDIDRTNIKKGLNEQHVNPWTTTLADLKNINNSMKIEKNNKSNEQVKNTS 448

Qy        709 I-RYVNEMDTSNKNNDNIFFDAISC--DNNMYPNITNNNNNNNNNNNNNNNNIDVENYNN 765

Db        449 VSKSCDIIKPSKFNKKNLFEQRLQSVEGKNFFEGRSQNLEGRSNFDERSQIVEQRRNFDD 508

Qy        766 RDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKN--NQDGNVNYSMNFCHYN--LN 821

Db        509 RD--QNIMDRKNFDERNQQVNDRRNFDERNQQVNDRRNFDDRDQNVMDRRNFDERNQQVN 566

Qy        822 DKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKK----KTNFYNNNNIVIVNNNMGNN 877

Db        567 DRRNFDERNQQVNDRRNFDDRDQNVMDRRNFDERNQQVNDRRNFDERNQQVNDRRNFDDR 626

Qy        878 NSPRM-KYGLCGSHTSIDNMKNNEMKNNEMKDN---EMKDNHIKSNNNNSSSSSSSNNNI 933

Db        627 DQNVMDRRNFDERNQQVNDRRNFDERNQQVNDRRNFDERNQHV----NDRRNFDERNQNV 682

Qy        934 YNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNI---YQNDDIIYTINSLNDYMSNTLL 990

Db        683 NDRRNFDERNQN---VNDRRNFDERNQQVNDRRNFDERYQN-------------VNERR 725

Qy        991 HFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDD-----NDNNNVDNNNKN-NVDNN 1044

Db        726 NFDERNQQVNDRRNFDERNQHVNERYQNVNDRRNFDERNQQVNDRRNFDERNQHVNERRN 785

Qy       1045 VD--NNNVDNNVDN-NDKNNVDNNN 1066

Db        786 FDERNQHVNERYQNVNDRRNFDERN 810
```

RESULT 34
US-09-820-843A-107
; Sequence 107, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTE
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: hypothetical protein
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: gi|3845248
US-09-820-843A-107

  Query Match             5.1%;  Score 348.5;  DB 11;  Length 665;
  Best Local Similarity   21.2%;  Pred. No. 2.8e-16;
  Matches  159;  Conservative  133;  Mismatches  310;  Indels  147;  Gaps   29;

Qy        421 SSLNDIDK---IKKNKKKG----LINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNN 473
                :| |||  :    ||| :|         |   |          | |:   |   |||   | :|: :|   :
Db         27 NSYNDIVEDINIKKEEKSSSEPPFIPIK------NKIDNVHTKNNNQYNLHNNKSNKTHL 80

Qy        474 NNNNDSSSMVNNMINHMIN-NNINNNINVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDF 532
                     :| : |   ||   :: :| :: ||  |       ::||||:     : : |      : ||
Db         81 TYGTHTSFLQNCTINDCVDVDNKDSEIN-NITKEKDDNNNNNGTKQIEEKNK--INKSDL 137

Qy        533 SSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTM 592
                   :::|       |:: |    ::|    :|        : |: :|   : |:      |          |::
Db        138 HRQNELNLQSGKNEQDINKNEKGKQD----ISNSNAENKKDVKEGVKELEEKKKEEKIS- 192

Qy        593 FSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDIS--TK 650
                 |   |  ::|    ||: :|:|:|      |          |                |:   |   |     |
Db        193 ---DDHKVEENK--KSDDHKVEENKKSDDHKVEEN------------KKSDDHKIEEVKK 235

Qy        651 DENHNDNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSISDVQIR 710
                 | |  :::                    :         :   | |  :        |    |  ||| :
Db        236 VEEHEEDE------------------EEDKKEKKSENKNKDENKDENDEDNDEISD-EDE 276

Qy        711 YVNEMDTSNKNNDNIFFDAISCDNN---------MYPNITNNNNNNNNNNNNNNNNIDVE 761
                  ::::       ||:|   |     |                :       ::    |   |   :      : | |
Db        277 VDDDVEEDKNENDDIDDDKKETDKTHLEEEENEIIEKEFSDKKKNGKNKDTKKEKSKDTE 336

Qy        762 NYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQ-DGNVNYSMNFCHYNL 820
                  ::|          |                : :   |:       :: :    |    : |  : ::
Db        337 KEKSKDIEKEKSKDKEKEKSKDKEKEKGKDKEKEKSKDIEKEKEKDKDI----------- 385

Qy        821 NDKNYLIDLNNKEQKDKNIH---GCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNN 877
                 :|      |      :::|||:|           |       ::|: :  ||||     |       ::::   :
Db        386 -EKEKSKDTAKEKEKDKDIEKEKSKDMEKLKNKQNDEKKKDDNEKKKNDKQDIHDDNDDE 444

Qy        878 NSPRMKYGLCGSHTSIDNMKNNEMKNNEMKD--NEMKDNHIKSNNNNSSSSSSSNNNIYN 935
                |                   :: ::  |: : :|  |:  | |:    |  |  | :  | : |  |  |
Db        445 N-------------DMEEIEENDDEEDEDEDMENKKKKKKGKNGNENGNENGSENGNENG 491

Qy        936 NINDDDTFQNDYCHNDNTFTIRRKN-NTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKE 994
                | |  ::      :      |:|           :| |  |   |       :  |    |     | :: :
Db        492 NENGNENENKNESENENENENENENGNENENEKENEKDKNIKEIENVTN---------AN 542

Qy        995 KYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVD--NNN-------KNNVDNNV 1045
                |   |   :: |  :|                 |      :  |  |::|  ||::|   ||:            |:|: |
Db        543 KENYEKINKNSEI--------TITKSNIDIYNNNRNNDIDKVNNHIFTNQQKKHNLHNEQ 594

Qy       1046 DNNNVDNNVDNNDKNNV-DNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLN 1104
                :    |     ||   |  ||:  |           :::    :||       ||                ::|||::|  ||
Db        595 NKFNETLNVSTNHKNHYEEKKKYESNMFNVDKRMHKNLT----------SMDTILHN-LN 643

Qy       1105 SLDISSMYDDTKEILNNILLSKYKAEKDN 1133
                  |    |  :  |   |  :||:              |  :|   |
Db        644 --DKLSHHKDLKNVLND------KKKKKN 664

RESULT 35
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT: Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
```

```
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12996
;   LENGTH: 6281
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12996

  Query Match             5.0%;  Score 342.5;  DB 9;  Length 6281;
  Best Local Similarity   18.9%;  Pred. No. 1.5e-14;
  Matches  237;  Conservative  188;  Mismatches  484;  Indels  343;  Gaps   55;

Qy        202 KIGVN-SQNDESSRSHAILNID-LKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTD 259
                | |:| ::|   :::: |   |:: |   :   |          : || ||          ||:    |:
Db         89 KNGLNGAENLRNAQNTAKQNLNTLSHLTNNQKSAISSQIDRAGH------VSEVTATKNA 142

Qy        260 GANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF---VGKSKSIM--------- 307
                  :|   :   |::  |    ::  |    :  | |::   |   ||   :    | ::::|:
Db        143 ATELNTQMGNLEQAIHDQNTVKQSVKFTDAD--KAKRDAYTNAVSRAEAILNKTQGANTS 200

Qy        308 -------IANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINE----------EDDTNTE 350
                           |  |:|               :     ||    ||   |  ||                :  |
Db        201 KQDVEAAIQNVSSAKNALNGDQNVTNAKNAAKNALNNLTSINNAQKRDLTTKIDQATTVA 260

Qy        351 RISILDSKGSEMNA--SSIEN-VVIKSNHLLSNNNN---------------------NKI 386
                 :  :  :   :::|    ::::| :  |:|  |  |  |  :                           ||
Db        261 GVEAVSNTSTQLNTAMANLQNGINDKTNTLASENYHDADSDKKTAYTQAVTNAENILNK- 319

Qy        387 NRGKINDKIERNNILK-----------NKSFDKPREGFTSTFGKYSSLNDI--DKIKKNK 433
                |  |    ||      |  |                |  : :: :        :|          |        ||:|:
Db        320 NSGSNLDKTAVENALSQVANAKGALNGNHNLEQAKSNANTTINGLQHLTTAQKDKLKQQV 379

Qy        434 KK-----GLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNN---------DNNNNNNNDSS 480
                ::       |:.  ||:      ||:|             |:    ||      |            :   |    |  |::
Db        380 QQAQNVAGVDTVKSSA---NTLNGAMGTLRNSIQDNTATKNGQNYLDATERNKTNYNNAV 436

Qy        481 SMVNNMINHMINNNIN-NNIN--VNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDD 537
                    | :||    |  |:: |  ||                :     : :||    ||        |:   |   |:
Db        437 DSANGVINATSNPNMDANAINQIATQVTSTKNALDGTHN--LTQAKQTATNAIDGAT--- 491

Qy        538 MNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSV----GSKLTMF 593
                  :||     |          |    :         :: |  :|           |:      :     :|  :       :|  |
Db        492 ---NLNKAQKDAL--KAQVTSAQRVANVTSIQQTANELNTAMGQLQHGIDDENATKQTQK 546

Qy        594 SYDPQKNKDNTF------FKSNINKME-DNTPKDILYESRNVSNMNGNVLLGLNKNTHHD 646
                 | :::|     :           |: :||       |: |                                 :::
Db        547 YRDAEQSKKTAYDQAVAAAKAILNKQTGSNSDK-----------------AAVDRALQQV 589

Qy        647 ISTKDENHNDNKI---------NNGVINIINNSNVNSINN-----SNMNSINNSNMNSNS 692
                 ||||   :  |  |:                 |  |  :|  |  |:       :          :  ::  :|        |:|:
```

```
Db        590 TSTKDALNGDAKLAEAKAAAKQNLGTLNHITNAQRTDLEGQINQATTVDGVNTVKTNANT 649

Qy        693 IYKSNYNSNQSIS--DVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNN- 749

Db        650 LDGAMNSLQGSINDKDATLRNQNYLD-ADESKRNAYTQAVTAAEGILNKQTGGNTSKADV 708

Qy        750 --------------NNNNNNNNIDVENYNNRDGTNNSMKLYAYN-SHNLFQPDN------ 788

Db        709 DNALNAVTRAKAALNGADNLRNAKTSATNTIDGLPNLTQLQKDNLKHQVEQAQNVAGVNG 768

Qy        789 ---NKNTSN---------IQNINTNKNNQ------DGNVNYSMNFCHYN--LNDKNYLID 828

Db        769 VKDKGNTLNTAMGALRTSIQNDNTTKTSQNYLDASDSNKN------NYNTAVNNANGVIN 822

Qy        829 LNNKEQKDKN---------------IHGCDN---------NIIQNRNDFEKKKKTNF--- 861

Db        823 ATNNPNMDANAINGMANQVNTTKAALNGAQNLAQAKTNATNTINNAHDLNQKQKDALKTQ 882

Qy        862 YNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNN 921

Db        883 VNNAQRVSDANNV-QHTATELNSAMTALKAAIADKERTKASGNYVNADQEKRQAYDSKVT 941

Qy        922 NSSSSSSSNNNIYNNINDDDTFQNDY------CHNDNTFTIRRKNNTNINSNIYQNDDII 975

Db        942 NAENIISGTPNATLTVNDVNSAASQVNAAKTALNGDNNLRVAKEHANNTIDGLAQ----- 996

Qy        976 YTINSLNDYMSNTLLHFKEKYT----YPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDN- 1030

Db        997 -----LNNAQKAKLKEQVQSATTLDGVQTVKNSSQTLNTAMKGLRDSIANEATIKAGQNY 1051

Qy       1031 NNVDNNNKNNVD----------NNVDNNNVDNN-----------------------VDN 1056

Db       1052 TDASPNNRNEYDSAVTAAKAIINQTSNPTMEPNTITQVTSQVTTKEQALNGARNLAQAKT 1111

Qy       1057 NDKNNVDN----NNVDND--DDDVD---------------------FHNIKN-FNNNEYL 1088

Db       1112 TAKNNLNNLTSINNAQKDALTRSIDGATTVAGVNQETAKATELNNAMHSLQNGINDETQT 1171

Qy       1089 SYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSL 1148

Db       1172 KQTQKYLDAEPSK-------KSAYDQAVNAAKAILT---KASGQNVDKAAVEQALQNVN- 1220

Qy       1149 EEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDL-VMCHICN 1199

Db       1221 -----------------STKTAL----NGDAKLNEAKAAAKQTLGTLTHINN 1251
```

RESULT 36
US-09-925-300-1228
; Sequence 1228, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1228
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (147)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (435)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1228

Query Match 4.9%; Score 336.5; DB 10; Length 460;
Best Local Similarity 28.7%; Pred. No. 1.2e-15;
Matches 100; Conservative 71; Mismatches 145; Indels 33; Gaps 10;

```
Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFD 62

Db         20 ARVRVAVRLRPFVDGTAGASDPPCVRGMDSCSLEIANWRNHQETLKY----QF--DAFYG 73

Qy         63 DTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY 122

Db         74 EXSTQQDIYAGSVQPILRHLLE-GQNASVLAYGPTGAGKTHTMLGS-----PEQPGVIPR 127

Qy        123 AAGDIFTFLNIYDKDNTKG------IFISFYEIYCGKLYDLLQKRK-MVAALENGKKEVV 175

Db        128 ALMDL---LQLTREEGAEGRPWXLSVTMSYLEIYQEKVLDLLDPASGDLVIREDCRGNIL 184

Qy        176 VKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSL--- 232

Db        185 IPGLSQKPISSFADFERHFLPASRNRTVGATRLNQRSSRSHAVL---LVKVDQRERLAPF 241

Qy        233 ----GKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRD 288

Db        242 RQREGKLYLIDLAGSEDNRRTGNKGLRLKESGA-INTSLFVLGKVVDALNQGLPRVPYRD 300

Qy        289 SELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNK 337

Db        301 SKLTRLLQDSLGGSAHSILIANIAPERRFYLDTVSALNFAARSKEVINR 349
```

RESULT 37
US-10-171-311-196
; Sequence 196, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 196
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-196

Query Match 4.9%; Score 333.5; DB 15; Length 890;
Best Local Similarity 25.7%; Pred. No. 4.7e-15;
Matches 117; Conservative 64; Mismatches 128; Indels 147; Gaps 15;

```
Qy          4 KIKVVVRKRPL--SELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIER------HEF 55
              |:|| :| |||  |||| :::|   : ::|  || :  |:     : |  ||      | |
Db         64 KVKVYLRVRPLLPSELE-RQEDQGCVRIENVETLVLQAPKDSFAL-KSNERGIGQATHRF 121

Qy         56 IVDKVFDDTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSD 115
                ::|   |   : :  |:| :: |: : |      : || |  ||||:|: |:  |
Db        122 TFSQIFGPEVGQASFFNLTVKEMVKDVLK-GQNWLIYTYGVTNSGKTHTIQGTIKDG--- 177

Qy        116 TPGIFQYAAGDIFTFL-------------------------------------------- 131
                ||   :   ||  |
Db        178 --GILPRSLALIFNSLQGQLHPTPDLKPLLSNEVIWLDSKQIRQEEMKKLSLLNGGLQEE 235

Qy        132 --------NIY---------------------------DKDNTK--------------- 140
                      ::|                           | |
Db        236 ELSTSLKRSVYIESRIGTSTSFDSGIAGLSSISQCTSSSQLDETSHRWAQPDTAPLPVPA 295

Qy        141 ----GIFISFYEIYCGKLYDLLQ------KRKMVAALENGKKEVVVKDLKILRVLTKEEL 190
                  |:|||:|||  |||||:      ||: :   |:       |||  :  |    ||
Db        296 NIRFSIWISFFEIYNELLYDLLEPPSQQRKRQTLRLCEDQNGNPYVKDLNWIHVQDAEE- 354

Qy        191 ILKMIDGVLLRKIGVNSQ-------NDESSRSHAILNIDLKDINKN----TSLGKIAFID 239
                   | |:|  :|          |  |||||:|:| :  :          : ::: |
Db        355 ------AWKLLKVGRKNQSFASTHLNQNSSRSHSIFSIRILHLQGEGDIVPKISELSLCD 408

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAM------DSDKNHIPFRDSELTK 293
              ||||||  |    :: :    :  |.|| ||   |  || |:       | :| :|||||:|:|:
Db        409 LAGSERCKD--QKSGERLKEAGNINTSLHTLGRCIAALRQNQQNRSKQNLVPFRDSKLTR 466

Qy        294 VLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSS 329
              | :   | |: :| || |::|  |   ::||:  ::|:
Db        467 VFQGFFTGRGRSCMIVNVNPCASTYDETLHVAKFSA 502
```

RESULT 38
US-09-351-794A-2
; Sequence 2, Application US/09351794A
; Patent No. US20020042382A1
; GENERAL INFORMATION:
; APPLICANT: DUFFY, PATRICK E.
; APPLICANT: OCKENHOUSE, CHRISTIAN F.
; TITLE OF INVENTION: SEQUESTRIN
; FILE REFERENCE: 38644-175519
; CURRENT APPLICATION NUMBER: US/09/351,794A
; CURRENT FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: 08/559,896
; PRIOR FILING DATE: 1995-11-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-351-794A-2

```
  Query Match             4.9%;  Score 333;  DB 9;  Length 652;
  Best Local Similarity   21.5%;  Pred. No. 3.4e-15;
  Matches  176;  Conservative  132;  Mismatches  283;  Indels  226;  Gaps   38;

Qy        335 KNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDK 394
              | ::|  |    |||        :  :  | |:  |    | :|:|  :        | :: ||
Db          6 KKENTNTNTNTSTNTSA-----NTNTSTNTSANTNTSTKESHILDESKLETFYRDEL-DK 59

Qy        395 IERNNI-------LKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYND 447
               : :  |         :  || |:   | |   | :::|: |::   :  |:  |
Db         60 MGKEEIETYFKGNIDKKSLDE--------FHKI-LLEELNKMDKDELYEM--YREEL--- 105

Qy        448 NTINKKHNNNNNNNNDNNDNNNDNNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNN 507
              | | ::   | |    |     :: || |::    :    :    :     |        | :
Db        106 NRIEQQKIRNMNKQQINKTYKDEINNMNSDQVDKIHREELEKIEKEKI--------NKMD 157

Qy        508 NNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSS 567
                :  :  :   | : :     |         ::|::   | | |  :    || ::  | ||:
Db        158 KDEIDKIYREELDKMD---RDAIYSMYIEDIS---NKNIKDLI---KNEKETNKDKNKKK 208

Qy        568 CDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNT--FFKSNINKME----DNTPKDI 621
                 :|  |||  :: :               |     :||:   :        ||:    |    :|
Db        209 DIDINKKKKDIDID---------VDIDKDIHKDHVEELYGEVKNKLSDEELDRMDRDA 258

Qy        622 LYE--SRNVSNMNGNVL--LGLNKNTHHDISTKDENHND--NKINNGVINIINNSNVNSI 675
              ||      :  || :  |   : |  :      |   |:: |  :  :||    || ::   :: |
Db        259 LYRVYLEELDRMNRDELYRVYLEELEKIDKEEKEKIHREKLHKIEKEKINKMDKDQIDKI 318

Qy        676 NNSNMNSINNSNMNS------NSIYKSNYNSNQSISDVQIRYVNEMD------------ 716
                 :| :::  :              | |      |   :  ::     |   |:|
Db        319 YEEELNKMDSDEIQHVRRAILEDIQKEKI-QNLELEEIDRLYKEELDRMDREARYEIPMR 377

Qy        717 -TSNKNNDNIFFDAISCDNNMYPNITNNN-----NNNNNNNNNNNNIDVENYNNRDGTN 770
                 |    |||     |  ::|        | |         :||::|||||||  |||  |||:    |
Db        378 NLSRNEKDNIIHRNIKNESNQKNKKENVNVFIIHDNNDSNNNNNNNNRDVNNLNNKHTNN 437

Qy        771 NSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLN 830
              |       ||             :| :      ::|:: :|  :             :  |   :||
Db        438 N------YN-------ENVEVELVVRNLDKDKGAK--------------IED---IIDYF 467

Qy        831 NKE-QKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIV-IVNNNMGNNNSPRMKYGLCG 888
              ||| :||||:                          | :|||  :|: :| :|:|
Db        468 NKEIKKDKNV-----------------------NVSNIVNFLNSKVGKDNTP-------- 496

Qy        889 SHTSIDNMKNNEM----KNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQ 944
                      | :  | |::        || ::      :::::||:        ::       ||
Db        497 ----IQHKKENQVDVVRKNIQI----IQEDNIKNKGQKDNTEMLDNN------------- 535

Qy        945 NDYCHNDNTFTIRRKNNTNINSNIYQNDDI--IYTINSLNDYMSNTLLHFKEKYTYPTLS 1002
                                 |  |||   :|   |||   :   |  |: |                     :
Db        536 --------------KEITNI--DIKNVDDIKNVGDIKSVGD-----------------IK 562

Qy       1003 TNEDIYN----KEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNND 1058
               : :||  |        |  ::|   |:    |    |    |   ||   ||    : :|   |  |||      :
Db        563 SVDDINNVDGIKNVDG--IKNVDGIKNVDGINNVGDINNAGDTNNAGDINNV-----GDI 615

Qy       1059 KNNVDNNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNV 1095
                   |:||    ||:: |:          |: |        ::     |:  |
Db        616 NNSVDIYNVEHIDEAEKKPNLDNPKKFDWTQVFKDKV 652
```

RESULT 39
US-09-801-368-392
; Sequence 392, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 392
; LENGTH: 1093
; TYPE: PRT

```
;  ORGANISM: Saccharomyces cerevisiae
US-09-801-368-392

  Query Match             4.8%;  Score 327;  DB 10;  Length 1093;
  Best Local Similarity   18.7%;  Pred. No. 1.8e-14;
  Matches  232;  Conservative  184;  Mismatches  381;  Indels  446;  Gaps   55;

Qy        303 SKSIMIANIS--PTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDD------------T 347
Db         23 SKSVLLAPHSNHPVIEIATYS-ETDVYECYIRGFETKIVMRRTKDDWINITQVFKIAQFS 81

Qy        348 NTERISILDSKGSEMNASSIEN---------VVIKSNHLLSNNNNNKINRGKIND----- 393
Db         82 KTKRTKILEKESNDMQHEKVQGGYGRFQGTWIPLDSAKFL-------VNKYEIIDPVVNS 134

Qy        394 ------------KIERNNILKN----------KSFDK-PRE--GFTSTFGKYSSLNDID 427
Db        135 ILTFQFDPNNPPPKRSKNSILRKTSPGTKITSPSSYNKTPRKKNSSSSTSATTTAAN--- 191

Qy        428 KIKKNKKKGLINY--KSTLYN-----------DNTINKKHNNNN----NNNND------N 464
Db        192 --KKGKKNASINQPNPSPLQNLVFQTPQQFQVNSSMNIMNNNDNHTTMNFNNDTRHNLIN 249

Qy        465 NDNNNDN------------------------------------------NNNNNNDSSS 481
Db        250 NISNNSNQSTIIQQQKSIHENSFNNNYSATQKPLQFFPIPTNLQNKNVALNNPNNNDSNS 309

Qy        482 MVNNMINHMINNNINNNINVNNNNNNNNNN---------------NSHNNHLPQPNYAF 526
Db        310 ---------YSHNIDNVINSSNNNNNGNNNNLIIVPDGPMQSQQQQHHHEYL------- 353

Qy        527 TDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSV 586
Db        354 TNNFNHSMMDSIT---NGNSKK---RRKKL-------NQSNEQQFYNQQEK---IQRH-- 395

Qy        587 GSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHD 646
Db        396 ------FKLMKQPLLWQSFQNPNDHHNE--------YCDSNGSNNNNNTV----ASNGSS 437

Qy        647 ISTKDENHNDNKINNGVINII-----NNSNVNSINN--------SNMNSINNSNMNSN-- 691
Db        438 IEVFSSNENDNSMNMSSRSMTPFSAGNTSSQNKLENKMTDQEYKQTILTILSSERSSDVD 497

Qy        692 -----SIYKS--NYNSNQSISD------------VQIRYVNEMDTSNKNNDNIFFDAISC 732
Db        498 QALLATLYPAPKNFNINFEIDDQGHTPLHWATAMANIPLIKMLITLNAN-------ALQC 550

Qy        733 DNNMYPNITNNNNNNNNNNNNNNNNI-----------DVENYNNR-----------DGTN 770
Db        551 NKLGFNCITKSIFYNNCYKENAFDEIISILKICLITPDV---NGRLPFHYLIELSVNKSK 607

Qy        771 NSMKLYAYNSH---NLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLI 827
Db        608 NPMIIKSYMDSIILSLGQQD-----YNLLKICLNYQDNIGNTPLHLSALNLNFEVYNRLV 662

Qy        828 DLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNS-------- 879
Db        663 YL-----------GASTDIL----NLDNESPASIMNKFNTPAGGSNSRNNNTKADRKLAR 707

Qy        880 --PRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSNNNIYNNI 937
Db        708 NLPQKNY-------------YQQQQQQQQPQNNVKIPKIIKTQHPDKEDSTADVNIAK-- 752

Qy        938 NDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYT 997
Db        753 TDSEVNESQYLHSN------QPNSTNMNT-------IMEDLSNINSFVTSSVIKDIKSTP 799

Qy        998 YPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDN---------- 1047
Db        800 SKILENSPILYRRRSQSIS---DEKEKAKDNEN---QVEKKKDPLNSVKTAMPSLESPSS 853

Qy       1048 --------------------NNVDNNVDN------NDKNNVDNNNVDNDDDDVDFHNIKN 1081
Db        854 LLPIQMSPLGKYSKPLSQQINKLNTKVSSLQRIMGEEIKNLDNEVVETES--------SI 905

Qy       1082 FNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINE 1141
Db        906 SNNKKRLITIAHQIEDAFDSVSNKTPINSISDLQSRIKET--SSKLNSEKQNFI------ 957

Qy       1142 DIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQIN------------------N 1183
Db        958 --------------QSLEKSQALKLATIVQDEESKVDMNTNSSSHPEKQEDEEPIPKSTS 1003

Qy       1184 ETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVV 1243
Db       1004 ETSSPK---------NTKADAKFS------------NTV--QESYDVNETLRLATELTIL 1040

Qy       1244 EYQNKSANSVLLNVSSNNGDIILLNK-----KLVQDNIKNSMD 1281
Db       1041 QFKRRMTTLKISEAKSKINSSVKLDKYRNLIGITIENIDSKLD 1083


RESULT 40
US-10-342-224-40
; Sequence 40, Application US/10342224
; Publication No. US20030162294A1
; GENERAL INFORMATION:
;  APPLICANT: Nathalie Verbruggen
;  TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress
;  FILE REFERENCE: CNN-012US
;  CURRENT APPLICATION NUMBER: US/10/342,224
;  CURRENT FILING DATE: . 2003-01-13
;  PRIOR APPLICATION NUMBER: US/09/762,154
;  PRIOR FILING DATE: 2002-02-02
;  PRIOR APPLICATION NUMBER: EP 98202634.6
;  PRIOR FILING DATE: 1998-08-04
;  NUMBER OF SEQ ID NOS: 123
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40
;   LENGTH: 959
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-342-224-40

  Query Match             4.8%;  Score 325.5;  DB 12;  Length 959;
  Best Local Similarity   23.8%;  Pred. No. 1.9e-14;
  Matches  176;  Conservative  118;  Mismatches  305;  Indels  141;  Gaps   33;

Qy          5 IKVVVRKRPLSELEKKKKDSDIITVK-NNCTLYIDEPRYKVDMTKYIERHEFIVDKV-FD 62
Db         49 VEVIGRIRDYP--DRKEKSPSILQVNTDNQTV-----RVRAD----VGYRDFTLDGVSFS 97

Qy         63 DTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY 122
Db         98 EQEGLEEFYKKFIEERIKGV-KVGNKCTIMMYGPTGAGKSHTM-----FGCGKEPGIVYR 151

Qy        123 AAGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKM--------VAALENGKKEV 174
Db        152 SLRDI---LGDSDQDGVTFVQVTVLEVYNEEIYDLLSTNSSNNLGIGWPKGASTKVRLEV 208

Qy        175 VVKDLKILRVLTKEEL--ILKMIDGVLLRKIGVNSQ--NDESSRSHAILNIDLKDINKNT 230
Db        209 MGKKAKNASFISGTEAGKISKEIVKVEKRRI-VKSTLCNERSSRSHCIIILDVPTVG--- 264

Qy        231 SLGKIAFIDLAGSERGADTVSQNKQT----QTDGANINRSLLALKECIRAMDSDKNHIPF 286
Db        265 --GRLMLVDMAGSE----NIDQAGQTGFEAKMQTAKINQGNIALKRVVESIANGDSHVPF 318

Qy        287 RDSELTKVLRDIFV-GKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEE- 344
Db        319 RDSKLTMLLQDSFEDDKSKILMILCASPDPKEMHKTLCTLEYGAKAKCIVRGSHTPNKDK 378

Qy        345 ---DDTNT-----ERISILD-----------SKGSEMNAS-----SIENVVIKSNHLLSN 380
Db        379 YGGDESASAVILGSRIAAMDEFIIKLQSEKKQKEKERNEAQKQLKKKEEEVAALRSLLTQ 438
```

```
Qy    381 NNNNKINRGKINDKI-ERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLIN 439
                  |  :| :|: ||  :||::  ||  |       ::     ::  ::  :: ::
Db    439 REACATNEEEIKEKVNERTQLLKSE-LDKKLEECRRMAEEF-----VEMERRRMEERIVQ 492

Qy    440 YKSTL----YNDNTINKKHNNNNNNNNDNNDNNNDN-NNNNNNDSSSMVNNM-------- 486
            :  |         |  :   :|  :  |          :  ::|   || :|
Db    493 QQEELEMMRRRLEEIEVEFRRSNGGSVDETSGFAKRLRSLYSDDDPGMVKSMDLDMGDPE 552

Qy    487 --------INHMINNNINNNINVNNNNNNNNNNNNSHNNNHLPQ---PNYAFTDTSDFSSL 535
                  ::|  :|  |::|          |        | | |         :   :   :
Db    553 PVKQVWGAVSHQSSNTISSNF-------TNLLQPKPSENMLTQMYPDRVCLSTVFEEEEV 605

Qy    536 DDMNCHLNNNDKSIF------------LHKKNLRDNIKLKNRSS-----CDNIMNKKKN 577
            ::    :   ||||                | |:|  :    |  :|        ||     |
Db    606 EEEEEKVIVEDKSICLITTPMPSLNSEGLGKENCFNGADDKESASSRRLRIQNIFTLCGN 665

Qy    578 NLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLL 637
               |::|| | :    :       ||| ||  |  || |    ::              |: :
Db    666 QRELSQHS-GQEEDQANIASPDKKDNQFF-SITNKAEALAVEEA---------KENNISV 714

Qy    638 GLNKNTHHDISTKDENHNDN 657
               :|   ||  |  |   ||
Db    715 DQRENGQLDIYVKWETAADN 734

Search completed: October  2, 2003, 16:30:25
Job time : 97 secs
```

GenCore version 5.1.6


```
OM protein - protein search, using sw model

Run on:         October  2, 2003, 16:06:35 ; Search time 92 Seconds
                                             (without alignments)
                                             2222.172 Million cell updates/sec

Title:          US-10-006-780-2
Perfect score:  6842
Sequence:       1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters:      1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      A_Geneseq_19Jun03:*
                1:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                2:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                3:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                4:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                5:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                6:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                7:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                8:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                9:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                10:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                11:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                12:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                13:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                14:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                15:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                16:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                17:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                18:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                19:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                20:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                21:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                22:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                23:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
                24:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	940	13.7	590	12	AAR11019	Apo-lipoprotein E
2	786	11.5	790	21	AAG41923	Arabidopsis thalia
3	786	11.5	794	21	AAG41922	Arabidopsis thalia
4	786	11.5	814	21	AAG41921	Arabidopsis thalia
5	783	11.4	803	22	ABB58104	Drosophila melanog
6	778	11.4	1844	21	AAB18250	Plasmodium falcipa
7	768.5	11.2	363	23	AAU77186	Human KinI-3 DNA f
8	768.5	11.2	381	23	AAU77184	Human KinI-3 DNA f
9	768.5	11.2	1368	23	AAU77182	Human kinesin moto
10	758.5	11.1	343	23	AAU77185	Human KinI-3 DNA f
11	758.5	11.1	373	23	AAU77183	Human KinI-3 DNA f
12	758.5	11.1	730	21	AAY49949	Xenopus laevis kin
13	752.5	11.0	405	23	AAE14504	Human mitotic cent
14	752.5	11.0	434	23	AAE14503	Human mitotic cent
15	752.5	11.0	473	23	AAE14502	Human mitotic cent
16	752.5	11.0	723	23	AAE14505	Human mitotic cent
17	750.5	11.0	725	23	AAE14506	Human mitotic cent
18	748.5	10.9	1817	21	AAB18301	Plasmodium falcipa
19	745.5	10.9	362	23	AAE14501	Human mitotic cent
20	745.5	10.9	391	23	AAE14500	Human mitotic cent
21	745.5	10.9	430	23	AAE14499	Human mitotic cent
22	745	10.9	729	22	ABB59500	Drosophila melanog
23	733	10.7	728	22	AAM39258	Human polypeptide
24	732.5	10.7	679	20	AAW88456	Human kinesin-rela
25	732.5	10.7	690	22	AAM39257	Human polypeptide
26	732.5	10.7	694	22	AAM41043	Human polypeptide
27	732.5	10.7	694	22	AAM41044	Human polypeptide
28	698	10.2	.1436	21	AAB18199	Plasmodium falcipa
29	696	10.2	807	21	AAB18311	Plasmodium falcipa
30	686.5	10.0	626	22	ABB59733	Drosophila melanog
31	676	9.9	2150	23	AAO22566	Wooden leg (WOL) g
32	672	9.8	410	21	AAB56496	Human prostate can
33	663	9.7	604	23	ABB05593	Mitotic centromere
34	604.5	8.8	537	7	AAP60452	Sequence of the As
35	596	8.7	1331	23	AAE29358	Plasmodium falcipa
36	587	8.6	2485	21	AAB18172	Plasmodium falcipa
37	579	8.5	2380	21	AAB18315	Plasmodium falcipa
38	577.5	8.4	2500	21	AAB18272	Plasmodium falcipa
39	576.5	8.4	342	17	AAW03625	Human lutropin-cho
40	569.5	8.3	2010	21	AAB18218	Plasmodium falcipa
41	563.5	8.2	3973	21	AAB18253	Plasmodium falcipa
42	552	8.1	2539	21	AAB18198	Plasmodium falcipa
43	540.5	7.9	4134	20	AAY31946	Plasmodium falcipa
44	538.5	7.9	1073	22	AAG67417	Amino acid sequenc
45	538.5	7.9	1073	24	ABG72692	Yeast cut7, a homo

ALIGNMENTS

```
RESULT 1
AAR11019
ID   AAR11019 standard; Protein; 590 AA.
XX
AC   AAR11019;
XX
DT   25-MAR-2003  (updated)
DT   17-MAY-1991  (first entry)
XX
DE   Apo-lipoprotein E induced cell differentiation copying polypeptide (1).
XX
KW   Apo-lipoprotein E; cell differentiation; hypercholestrolaemia;
KW   androgen; diagnosis.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..20
FT                   /label= leader_segment
FT                   /note= "may be replaced by NH2"
FT   Region          21..40
FT                   /label= first_spacing_segment
FT   Region          /note= "may be omitted"
FT   Region          41..55
FT                   /label= p(141-155)
FT   Region          56..76
FT                   /label= second_spacing_segment
FT   Region          21..130
FT                   /label= repeated_segment
FT                   /note= "present at least twice"
XX
PN   WO9102751-A.
```

```
XX
PD   07-MAR-1991.
XX
PF   17-AUG-1990;   90WO-US04651.
XX
PR   18-AUG-1989;   89US-0395732.
PR   26-FEB-1990;   90US-0485158.
PR   18-JUN-1990;   90US-0540363.
XX
PA   (SCRI ) SCRIPPS CLINIC & RES FOUND.
XX
PI   Dyer CA,  Curtiss LK,  Smith R;
XX
DR   WPI; 1991-087246/12.
XX
PT   Polypeptide(s) copying apo-lipoprotein E induced cell
PT   differentiation - useful for treating hypercholestrolaemia,
PT   inhibiting ovarian androgen secretion and for prodn. of diagnostic
PT   systems.
XX
PS   Claim 2; Page 98; 123pp; English.
XX
CC   A polypeptide capable of inhibiting lymphocyte proliferation and/or
CC   ovarian androgen secretion comprises a plurality of segments, as
CC   indicated in the features.  The segment may be repeated 2-10 times.
CC   In this example the segment is repeated 10 times.  The N-terminal
CC   may comprise NH2 or a leader segment of 1-20 amino acids, the
CC   C-terminal comprises COOH or a tail segment of 1-20 amino acids.
CC   Between each segment is a first and second spacing segment of
CC   1-20 amino acids.  p(141-155) is a peptide obtained from apo E.
CC   See also AAR11107 and AAR11109.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence   590 AA;

  Query Match             13.7%;  Score 940;  DB 12;  Length 590;
  Best Local Similarity   32.4%;  Pred. No. 4.4e-45;
  Matches  235;  Conservative   99;  Mismatches  215;  Indels  176;  Gaps   18;

Qy        374 SNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNK 433
              :|:  :||||  |      |:      ||     |  :  |  |:             |  |  |  :  |
Db          9 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLRKLRKRL---------LRDADDLNNNN 59

Qy        434 KKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSM----------- 482
                          |:|   |   :||||||||:||:|||:||||||:::::
Db         60 ------------NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLRKLRKRLLRDA 107

Qy        483 --VNNMINHMINNNINNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNC 540
                :||  |:  |||  |||  |  ||||||||||::||:
Db        108 DDLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN--------------------- 145

Qy        541 HLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKN 600
                |||:     |  |:  |||      |  |  ::  :|    |      ||                      |
Db        146 --NNNNNLRKLRKRLLRDADDLNNNNNNNNNNNNNN--------------------N 182

Qy        601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 660
               :|      :|  |    :|               :  |  :|  |  |  |  |  |          :  |:|:|·  |
Db        183 NNN----NNNNNNNNN--------NNNNNNNNNNNLRKLRKRLLRDADDLNNNNNNNNNN 230

Qy        661 NGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSIDVQIRYVNEMDTSNK 720
              |         ||:|  |:  ||:|  |:  ||:|  |:|           |:|  ::    ::  |  :  :  |    |
Db        231 N------NNNNNNNNNNNNNNNNNNNNNNNN-------NNNNNLRKLRKRLLRDADDLNN 277

Qy        721 NNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNS 780
              ||:|              ||       |    ||||||||||||||        |  ||   :    ||   ||
Db        278 NNNN---------NNNNNNNNNNNNNNNNNNNNNNNN----NNNNNNNNNNLRKLRKRLL 324

Qy        781 HNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLI-----------D 828
               :    :||  |  :|    |  |  |  ||  :  |  |  :  |    :  |  |:  |    :              |
Db        325 RDADDLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLRKLRKRLLRDADD 384

Qy        829 LNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCG 888
              |||      :  |        ||     |  |:          |    ||||        |||    |||      :::   |
Db        385 LNNNNNNNNN-----NNNNNNNNNNNNNNNNNNNNNNNN----NNNNNNNNLRKLRKRLL- 434

Qy        889 SHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNN--------INDD 940
                          |::  ||      ||      :|       :|:    :||||::::::::|||    ||                    :  |
Db        435 --RDADDLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLRKLRKRLLRDA 492

Qy        941 DTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPT 1000
                |      |:  :|:|              ||  |  |:|        |::           |:  |:    :|  |        :::
Db        493 DDLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-----NNNNNNNNNNLRKLRKRL---- 543

Qy       1001 LSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKN 1060
                                                  |  |  |     ::|:|||  :|||  ||  :||  :|||  :||  :||:  |
Db        544 -----------------LRDADDLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 585

Qy       1061 NVDNN 1065
              |  :||
Db        586 NNNNN 590
```

```
RESULT 2
AAG41923
ID   AAG41923 standard; Protein; 790 AA.
XX
AC   AAG41923;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52219.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
```

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

```
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             11.5%;  Score 786;  DB 21;  Length 790;
  Best Local Similarity   34.8%;  Pred. No. 2.9e-36;
  Matches  210;  Conservative  103;  Mismatches  183;  Indels  108;  Gaps   21;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFD 62
               :||||||||||:: |  ||: |::|| :| :|  : ||| |||:| |:|:||| | | |
Db        188 AKIKVVVRKRPLNKKETAKKEEDVVTVSDN-SLTVHEPRVKVDLTAYVEKHEFCFDAVLD 246

Qy         63 DTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY 122
               : |  |  ||  ||:|:|  :::      :||||||||||||:||   :|          :
Db        247 EDVSNDEVYRATIEPIIPIIFQR-TKATCFAYGQTGSGKTFTM---KP--------LPIR 294

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
               | |:   |   :|      | :::|::||| |||:||| :|| :   |:|:::| :  |:
Db        295 AVEDLMRLLRQPVYSNQRFK-LWLSYFEIYGGKLFDLLSERKKLCMREDGRQQVCIVGLQ 353

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAIL------NIDLKDINKNTS-- 231
                  | :  :::   |: |   |  |     |:|||||||||       ::::|| :  :
Db        354 EYEV-SDVQIVKDFIEKGNAERSTGSTGANEESSRSHAILQLVVKKHVEVKDTRRRNNDS 412

Qy        232 -------LGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHI 284
                      :|||:||||||||||||||     ::||: :|| ||:||||||||||:|:|: ||
Db        413 NELPGKVVGKISFIDLAGSERGADTTDNDRQTRIEGAEINKSLLALKECIRALDNDQLHI 472

Qy        285 PFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVK------NFKNKS 338
               ||| |:||:|||| ||| |:::||: |||      || |||||||: |||        | |
Db        473 PFRGSKLTEVLRDSFVGNSRTVMISCISPNAGSCEHTLNTLRYADRVKSLSKSGNSKKDQ 532

Qy        339 TC-----INEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKIND 393
               |        :|::         :    :    |:|         |:: :     ||:: : |:  :  :
Db        533 TANSMPPVNKDPLLGPNDVEDVFEPPQEVNVPETRRRVVEKD---SNSSTSGIDFRQPTN 589

Qy        394 KIERNNILKNKSFDKPREGFTSTF-GKYSSLNDI-----------DKIKK---NKKKGLI 438
                   | : |   : | || |      |:| |   |   |:|              :|:||     : |||
Db        590 YREESGI-PSFSMDKGRSEPNSSFAGSTSQRNNISSYPQETSDREEKVKKVSPPRGKGL- 647

Qy        439 NYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNDSSSMVNNMINHMINNNINNN 498
                                                :  :  |  :  | ||                ::
Db        648 -------------------------REEKPDRPQNWSKRDVSS---------------SD 667

Qy        499 INVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRD 558
               |     |     |  :    |                 |  | ||  :|| :    |    :  |  |:|  : |
Db        668 IPTLTNFRQNASETASRQYETASRQYE-TDPSLDENLDAL---LEEEEALIAAHRKEIED 723

Qy        559 NIKL 562
                 :::
Db        724 TMEI 727
```

```
RESULT 3
AAG41922
ID   AAG41922 standard; Protein; 794 AA.
XX
AC   AAG41922;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52218.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
```

```
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             11.5%;  Score 786;  DB 21;  Length 794;
  Best Local Similarity   34.8%;  Pred. No. 3e-36;
  Matches  210;  Conservative  103;  Mismatches  183;  Indels  108;  Gaps   21;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFD 62
              :|||||||||||:: |  ||: |::|| :| :| : ||| |||:| |:|:|||  | | |
Db        192 AKIKVVVRKRPLNKKETAKKEEDVVTVSDN-SLTVHEPRVKVDLTAYVEKHEFCFDAVLD 250

Qy         63 DTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY 122
              : | |  ||  ||:|:|  :::     :||||||||||||:||   :|        :
Db        251 EDVSNDEVYRATIEPIIPIIFQR-TKATCFAYGQTGSGKTFTM---KP--------LPIR 298

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |  |:   |   :|     | :::|::||| |||:||| :|| :   |:|:::| :  |:
Db        299 AVEDLMRLLRQPVYSNQRFK-LWLSYFEIYGGKLFDLLSERKKLCMREDGRQQVCIVGLQ 357

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAIL------NIDLKDINKNTS-- 231
                 | :  :::   |: |   |  |    |:|||||||||      ::::||  :  :
Db        358 EYEV-SDVQIVKDFIEKGNAERSTGSTGANEESSRSHAILQLVVKKHVEVKDTRRRNNDS 416

Qy        232 -------LGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHI 284
                     :|||:|||||||||||||   ::||: :|| ||:|||||||||||:|:|: ||
Db        417 NELPGKVVGKISFIDLAGSERGADTTDNDRQTRIEGAEINKSLLALKECIRALDNDQLHI 476

Qy        285 PFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVK------NFKNKS 338
              ||| |:||:||||  ||| |:::||: |||    ||  |||||||: |||      | |
Db        477 PFRGSKLTEVLRDSFVGNSRTVMISCISPNAGSCEHTLNTLRYADRVKSLSKSGNSKKDQ 536

Qy        339 TC-----INEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKIND 393
              |      :|::       :  :     |:|       |:: :   ||:: : |:  :  :
Db        537 TANSMPPVNKDPLLGPNDVEDVFEPPQEVNVPETRRRVVEKD---SNSSTSGIDFRQPTN 593

Qy        394 KIERNNILKNKSFDKPREGFTSTF-GKYSSLNDI-----------DKIKK---NKKKGLI 438
                | : |  : | || |    |:| |  |  |:|           :|:||    : |||
Db        594 YREESGI-PSFSMDKGRSEPNSSFAGSTSQRNNISSYPQETSDREEKVKKVSPPRGKGL- 651

Qy        439 NYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNN 498
```

```
                                     : :   | :  | ||              ::
Db        652 ------------------------REEKPDRPQNWSKRDVSS---------------SD 671

Qy        499 INVNNNNNNNNNNNNSHNHHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRD 558
              |    |    | :   |          |  |||  :|| :   |    :  |  |:| : |
Db        672 IPTLTNFRQNASETASRQYETASRQYE-TDPSLDENLDAL---LEEEEALIAAHRKEIED 727

Qy        559 NIKL 562
               :::
Db        728 TMEI 731


RESULT 4
AAG41921
ID   AAG41921 standard; Protein; 814 AA.
XX
AC   AAG41921;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52217.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
```

```
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             11.5%;  Score 786;  DB 21;  Length 814;
  Best Local Similarity   34.8%;  Pred. No. 3e-36;
  Matches  210;  Conservative  103;  Mismatches  183;  Indels  108;  Gaps   21;

Qy          3 SKIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFD 62
Db        212 AKIKVVVRKRPLNKKETAKKEEDVVTVSDN-SLTVHEPRVKVDLTAYVEKHEFCFDAVLD 270

Qy         63 DTVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQY 122
Db        271 EDVSNDEVYRATIEPIIPIIFQR-TKATCFAYGQTGSGKTFTM---KP--------LPIR 318

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
Db        319 AVEDLMRLLRQPVYSNQRFK-LWLSYFEIYGGKLFDLLSERKKLCMREDGRQQVCIVGLQ 377

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAIL------NIDLKDINKNTS-- 231
Db        378 EYEV-SDVQIVKDFIEKGNAERSTGSTGANEESSRSHAILQLVVKKHVEVKDTRRRNNDS 436

Qy        232 -------LGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHI 284
Db        437 NELPGKVVGKISFIDLAGSERGADTTDNDRQTRIEGAEINKSLLALKECIRALDNDQLHI 496

Qy        285 PFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVK------NFKNKS 338
Db        497 PFRGSKLTEVLRDSFVGNSRTVMISCISPNAGSCEHTLNTLRYADRVKSLSKSGNSKKDQ 556

Qy        339 TC-----INEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKIND 393
Db        557 TANSMPPVNKDPLLGPNDVEDVFEPPQEVNVPETRRRVVEKD---SNSSTSGIDFRQPTN 613

Qy        394 KIERNNILKNKSFDKPREGFTSTF-GKYSSLNDI-----------DKIKK---NKKKGLI 438
Db        614 YREESGI-PSFSMDKGRSEPNSSFAGSTSQRNNISSYPQETSDREEKVKKVSPPRGKGL- 671

Qy        439 NYKSTLYNDNTINKKHNNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNN 498
Db        672 ------------------------REEKPDRPQNWSKRDVSS---------------SD 691

Qy        499 INVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRD 558
Db        692 IPTLTNFRQNASETASRQYETASRQYE-TDPSLDENLDAL---LEEEEALIAAHRKEIED 747

Qy        559 NIKL 562
Db        748 TMEI 751

RESULT 5
ABB58104
ID   ABB58104 standard; Protein; 803 AA.
XX
AC   ABB58104;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 1104.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
```

```
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02207.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 1104; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   803 AA;

  Query Match             11.4%;  Score 783;  DB 22;  Length 803;
  Best Local Similarity   38.8%;  Pred. No. 4.4e-36;
  Matches  193;  Conservative   78;  Mismatches  153;  Indels   74;  Gaps   13;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||:|  |   :|: |:|:|     | : ||| |||:||::| |:|  |  |:|
Db        276 QITVCVRKRPISRKEVNRKEIDVISVPRKDMLIVHEPRSKVDLTKFLENHKFRFDYAFND 335

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYG--QSDTPGIFQ 121
              | ||  ||: | |||:  ::| | : :|||||||||||:|| | :  |  |      ||:
Db        336 TCDNAMVYKYTAKPLVKTIFEGG-MATCFAYGQTGSGKTHTM-GGEFNGKVQDCKNGIYA 393

Qy        122 YAAGDIFTFLNI--YDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDL 179
               || |:|  ||:  |    |    :  ||:||| ||::|||  :: :  ||:||::| |  |
Db        394 MAAKDVFVTLNMPRYRAMNLV-VSASFFEIYSGKVFDLLSDKQKLRVLEDGKQQVQVVGL 452

Qy        180 KILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFI 238
               :|:    | :||:|  |    |   |   |  |  ||||||:  | |:          || :||
Db        453 -TEKVVDGVEEVLKLIQHGNAARTSGQTSANSNSSRSHAVFQIVLRPQGSTKIHGKFSFI 511

Qy        239 DLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDI 298
              ||||:|||  || | ::|:  :|| |:|||||||||||:        |:|||  |:||:|||
Db        512 DLAGNERGVDTSSADRQTRMEGAEINKSLLALKECIRALGKQSAHLPFRVSKLTQVLRDS 571

Qy        299 FVG-KSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDS 357
              |:| |||: ||| ||| :| || |||||||: |||     |     :       :|| | | |
Db        572 FIGEKSKTCMIAMISPGLSSCEHTLNTLRYADRVKELVVKD-IVEVCPGGDTEPIEITDD 630

Qy        358 KGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPREGFTSTF 417
               :  |          |:|    :| |   |::    ::
Db        631 EEEEE-----LNMVHPHSHQLHPNSHAPASQ---------------------------- 656

Qy        418 GKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNNNN 477
                            |::    ::    :                |||||||  : |  | |
Db        657 ------------SNNQRAPASHHSGAVI--------HNNNNNNNKNGNAGNMD------- 689

Qy        478 DSSSMVNNMINHMINNNI 495
                :|::::   |  ::: :
Db        690 --LAMLSSLSEHEMSDEL 705

RESULT 6
AAB18250
ID   AAB18250 standard; Protein; 1844 AA.
XX
AC   AAB18250;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:107.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 246-250; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   1844 AA;

  Query Match             11.4%;  Score 778;  DB 21;  Length 1844;
  Best Local Similarity   24.5%;  Pred. No. 2.2e-35;
  Matches  340;  Conservative  202;  Mismatches  439;  Indels  406;  Gaps   66;

Qy        205 VNSQNDESSRSHAILNI-DLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQT----- 258
              || :|| :   :  :|    : :|| ||:  : |:|:          :| ::  |  |
Db         37 VNRRNDMNYEMYNSINSGHMSNINNNTNNLQDAYINKELHYMNSDKINIAKNHQNVHMTA 96

Qy        259 -----DGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISP 313
                   |  |  | :::  |    |   |   |:|:  |  :|          :        : ::   | :
Db         97 TYNNMDKNNANNNIIHLNNHIN-MTDDQNY--FYNSTTNNKM-------NNTLKENNFNN 146

Qy        314 TISCCEQTLNTLRYSSRVKNFKNKSTCINEEDD-TNTERISILDSKGSEMNASSIEN--- 369
               ::       |:|   | :     || |  :      ||:|     |    |:  :::   :  | :  :
Db        147 NMN----TVNNSFYHNTDNNFLNFTRNQNEQDTYVNNNIINNFNNQNVDKNINNNNNNPN 202

Qy        370 ----VVIKSNHLLSNNNNN----KINRGKINDKIERNNILKNKSFDKPREGFTSTFGKYS 421
                  : | ||: :  |  |        |: || |   :| |  |  | :                |
Db        203 KNVESINKFNHIYNMQNFNHFIPNISNGK-NGNLENNASLSH-----------NVNSVS 249
```

```
Qy    422 SLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNN---NNND 478
Db    250 TISELHNFNYMNNMDLNNVDMNNMNMNNMNMNNMNMNNMNMNNMDMNNVNMNNMNMNNMD 309
Qy    479 SSSM-------VNNM-INHMINNNIN-NNINVNNNNNNNN-----------NNNNSHN-N 517
Db    310 VNSMNMNNMDNMNNMNMNNMDVNNMNMNNMDVNNMDNMNNMDNMNNMKNLSNFNNSYQYN 369
Qy    518 HLPQPN----------------YAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKN------ 555
Db    370 SIPQFNSSSRFNNITHFNNGISHNVNNVSNFSN----NAHLDNSNNMNRLNAVNNFGDIN 425
Qy    556 -----------LRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQ------ 598
Db    426 SFHDPLNEMQVLNKNVNMQNENHF-NVMNDEMKNYNNVK-----RINSISHIPYMNNLKN 479
Qy    599 --------KNKDNTFFKSNINKMEDNTPKDI---LYESRNVSNMNGNVLLG-LNKNTHHD 646
Db    480 YNEHTSMVKGKGNTNRKKSNNLKINNNPGSVNARAISENNQSTAHGNIPMGSVDKVIKHD 539
Qy    647 ISTKD-----ENHNDNKIN-----NGVINIINNSNVNSINN-------SNMNSINN-SNM 688
Db    540 RMDNDLKNINNMNNMNSMNNMNSMNNMNNMINMNNMNNMNNMINMNNMNNMNNMNNMSNM 599
Qy    689 NSNSIYKSNYNSNQSI---------SDVQIRYVNEMDTSNK----------------NND 723
Db    600 NNTSILN---NNNKKITKRGRAKKNSTININNINKMNSTNNKSSMINMNSVNNMNSVNNM 656
Qy    724 NIFFDAISCDN-NMYPNITNNNNNNNNNNNNNNNI-DVENYNNRDGTNNSMKLYAYNS- 780
Db    657 NSVKNMNSVNNMNSVNNINNVNNINNVNNINNVNNINNVNNINNVNNINNVNNINNVNNM 716
Qy    781 HNLF-----------QPDNNKNTSNIQNIN---------TNKNNQDGNVNYSMNFCHYNL 820
Db    717 NNMFNVNPQLNIMGIMKDINNNNITVSNKNKLMNNYINDNNIMNMEGSINETYNF-DGTL 775
Qy    821 NDKNYLIDLNNKEQKDKNIHGCDN------------------NIIQNRNDFEKKKKTNFY 862
Db    776 NNKN-VSNNNNNDIHDKGVLNTLNRSKSSSYIKPHRNLTLPINMYMNNTYMYNSKAYVNY 834
Qy    863 NNNNIVIV-----NNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIK 917
Db    835 ENQNWMAQQDCNDKNGLSLNEGPRYN----------DNNNNNNNNNNNNNNNNNNNN-N 883
Qy    918 SNNNNSSSSSSSNNNIYNNINDDDTFQNDY----CHNDNTFTIRRKNNTNIN-------S 966
Db    884 NNNNNNNNHSIINNNITQGIHRGSMINNQHFDELSYNPNGIFLEKNTITNYNEMLGSGYN 943
Qy    967 NIY---------------QNDDIIYTINSLNDYMSN---------------------TLL 990
Db    944 NMYDKNAVKGNMNLIGKHSNYDLMKNGNMINGYKLNMQEVQKGMNEVGKKRAPRRRAKTI 1003
Qy    991 HFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDN----------------NNVD 1034
Db   1004 NLKSINRFNTVSLG-------LKGNH-NMENMNNMDNMDNMNNMINVSNVNNVNNMNNVN 1055
Qy   1035 N-NNKNNVD--NNVDN-NNVD--NNVDN-NDKNNVDN----NNVDNDDDDVDFHNIKNFN 1083
Db   1056 NVNNMNNVNNMNNVNNMNNVNSVNNMNNTNNMNNMNNVNSVNNMNNTNNMNNMNNTNNMN 1115
Qy   1084 NNEYLSYFQKN-------VDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIK 1136
Db   1116 NMNRMNHIDNNMVTNLNYMDNKINNAGNNLN----GEFSKRILYN--------------- 1156
Qy   1137 KYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSD-----LRKD 1191
Db   1157 -------RSKSTENIKEMMQPHNEANNI----------------SGNNTSDSNNITLNKN 1193
Qy   1192 LVMCHICNNNP--DDQFH---FYAYSRLEKDIINLIMLRQIWCE--------SENLRLLY 1238
Db   1194 LVEMILRNNKPSIDKNIHESVNRSYTSFLMNIGSSYLKKKKTAEIKTGENNTENNKGIVN 1253
Qy   1239 QFLVVEYQNKSANSVLLNV------------SSNNGDIILLNKKLVQDNIKN-----SMD 1281
Db   1254 VNSQVEEKGKSENKSILEIKKEGQIKNVIYNNNNNNNNEKDQNADQYQDQNKNHKQDQRQD 1313
Qy   1282 HNNIHKK 1288
Db   1314 QNKSHKQ 1320
```

RESULT 7
AAU77186
ID AAU77186 standard; Protein; 363 AA.
XX
AC AAU77186;
XX
DT 02-JUL-2002 (first entry)
XX
DE Human KinI-3 DNA fragment with flanking vector sequences #4.
XX
KW Human; kinesin motor protein; KinI-3; microtubule; hyperplasia;
KW depolymerisation; cellular proliferation disorder; cancer; restenosis;
KW vascular malfunction; abnormal wound healing; cardiac hypertrophy; gout;
KW inflammatory disorder; immune disorder; rheumatoid arthritis; psoriasis;
KW diabetic retinopathy; neurological disorder; immunostimulant; vasotropic;
KW vesicular transport disorder; immunosuppressive; antiinflammatory;
KW cytostatic.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200226929-A2.
XX
PD 04-APR-2002.
XX
PF 28-SEP-2001; 2001WO-US30750.
XX
PR 29-SEP-2000; 2000US-0675227.
XX
PA (CYTO-) CYTOKINETICS INC.
XX
PI Beraud C, Guo J, Freedman R, Patel UA, Davies KA;
XX
DR WPI; 2002-352124/38.
DR N-PSDB; ABK49250.
XX
PT An novel isolated microtubule motor protein, KinI-3 used for preparing
PT compositions for the diagnosis, prevention or treatment of diseases
PT such as cancers, hyperplasias or restenosis -
XX
PS Disclosure; Fig 6; 68pp; English.
XX
CC The invention relates to a human kinesin motor protein (KinI-3), with
CC microtubule depolymerisation activity, and the polynucleotide encoding
CC it. The novel KinI-3 motor protein can be used to prepare compositions
CC for the diagnosis, prevention or treatment of cellular proliferation
CC disorders, cancers, hyperplasias, restenosis, vascular malfunctions,
CC abnormal wound healing, cardiac hypertrophy, inflammatory and immune
CC disorders (such as rheumatoid arthritis, gout, psoriasis and diabetic
CC retinopathy) neurological disorders and disorders of vesicular transport.
CC This sequence represents a human KinI-3 polypeptide fragment with
CC flanking vector sequences.
XX
SQ Sequence 363 AA;

Query Match 11.2%; Score 768.5; DB 23; Length 363;
Best Local Similarity 47.5%; Pred. No. 1.2e-35;
Matches 164; Conservative 58; Mismatches 116; Indels 7; Gaps 3;

```
Qy      4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db     12 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 71
Qy     64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
```

```
Db          72 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 125

Qy         124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
               | |||  | :          ::|||||||||:|||||  :||  :  |  |:  |   |  :   |: |:
Db         126 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 185

Qy         184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
                | : | |:  ::  |    |  |       | :||||||:: | :||  |  |   |:|:|||||||
Db         186 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 244

Qy         244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
               || ||     ::||: :||  ||:|||||||||||:|  :    |  |||  |:||:||:|  |:|  :
Db         245 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 304

Qy         304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
               |: ||||||:      |||||||: |||   |        |
Db         305 KTCMIANISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRN 349
```

RESULT 8
AAU77184
ID AAU77184 standard; Protein; 381 AA.
XX
AC AAU77184;
XX
DT 02-JUL-2002 (first entry)
XX
DE Human KinI-3 DNA fragment with flanking vector sequences #2.
XX
KW Human; kinesin motor protein; KinI-3; microtubule; hyperplasia;
KW depolymerisation; cellular proliferation disorder; cancer; restenosis;
KW vascular malfunction; abnormal wound healing; cardiac hypertrophy; gout;
KW inflammatory disorder; immune disorder; rheumatoid arthritis; psoriasis;
KW diabetic retinopathy; neurological disorder; immunostimulant; vasotropic;
KW vesicular transport disorder; immunosuppressive; antiinflammatory;
KW cytostatic.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200226929-A2.
XX
PD 04-APR-2002.
XX
PF 28-SEP-2001; 2001WO-US30750.
XX
PR 29-SEP-2000; 2000US-0675227.
XX
PA (CYTO-) CYTOKINETICS INC.
XX
PI Beraud C, Guo J, Freedman R, Patel UA, Davies KA;
XX
DR WPI; 2002-352124/38.
DR N-PSDB; ABK47999.
XX
PT An novel isolated microtubule motor protein, KinI-3 used for preparing
PT compositions for the diagnosis, prevention or treatment of diseases
PT such as cancers, hyperplasias or restenosis -
XX
PS Disclosure; Fig 4; 68pp; English.
XX
CC The invention relates to a human kinesin motor protein (KinI-3), with
CC microtubule depolymerisation activity, and the polynucleotide encoding
CC it. The novel KinI-3 motor protein can be used to prepare compositions
CC for the diagnosis, prevention or treatment of cellular proliferation
CC disorders, cancers, hyperplasias, restenosis, vascular malfunctions,
CC abnormal wound healing, cardiac hypertrophy, inflammatory and immune
CC disorders (such as rheumatoid arthritis, gout, psoriasis and diabetic
CC retinopathy) neurological disorders and disorders of vesicular transport.
CC This sequence represents a human KinI-3 polypeptide fragment with
CC flanking vector sequences.
XX
SQ Sequence 381 AA;

```
  Query Match             11.2%;  Score 768.5;  DB 23;  Length 381;
  Best Local Similarity   47.5%;  Pred. No. 1.3e-35;
  Matches  164;  Conservative   58;  Mismatches  116;  Indels    7;  Gaps    3;

Qy           4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
               ||:| ||||||    | :: : :|||::  || : | :   ||:|:||  :|  |   |:||  :
Db          30 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 89

Qy          64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                    |  ||  |   |||   ::  ||    :|||||||||:||||||:|:         :  ||::  |
Db          90 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 143

Qy         124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
               | |||  | :          ::|||||||||:|||||  :||  :  |  |:  |   |  :   |: |:
Db         144 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 203

Qy         184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
                | : | |:  ::  |    |  |       | :||||||:: | :||  |  |   |:|:|||||||
Db         204 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 262

Qy         244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
               || ||     ::||: :||  ||:|||||||||||:|  :    |  |||  |:||:||:|  |:|  :
Db         263 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 322

Qy         304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
               |: ||||||:      |||||||: |||   |        |
Db         323 KTCMIANISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRN 367
```

RESULT 9
AAU77182
ID AAU77182 standard; Protein; 1368 AA.
XX
AC AAU77182;
XX
DT 02-JUL-2002 (first entry)
XX
DE Human kinesin motor protein KinI-3.
XX
KW Human; kinesin motor protein; KinI-3; microtubule; hyperplasia;
KW depolymerisation; cellular proliferation disorder; cancer; restenosis;
KW vascular malfunction; abnormal wound healing; cardiac hypertrophy; gout;
KW inflammatory disorder; immune disorder; rheumatoid arthritis; psoriasis;
KW diabetic retinopathy; neurological disorder; immunostimulant; vasotropic;
KW vesicular transport disorder; immunosuppressive; antiinflammatory;
KW cytostatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 224..545
FT /label= Motor_domain
XX
PN WO200226929-A2.
XX
PD 04-APR-2002.
XX
PF 28-SEP-2001; 2001WO-US30750.
XX
PR 29-SEP-2000; 2000US-0675227.
XX
PA (CYTO-) CYTOKINETICS INC.
XX
PI Beraud C, Guo J, Freedman R, Patel UA, Davies KA;
XX
DR WPI; 2002-352124/38.
DR N-PSDB; ABK47997.
XX
PT An novel isolated microtubule motor protein, KinI-3 used for preparing
PT compositions for the diagnosis, prevention or treatment of diseases
PT such as cancers, hyperplasias or restenosis -
XX

```
PS   Claim 11; Fig 2; 68pp; English.
XX
CC   The invention relates to a human kinesin motor protein (KinI-3), with
CC   microtubule depolymerisation activity, and the polynucleotide encoding
CC   it. The novel KinI-3 motor protein can be used to prepare compositions
CC   for the diagnosis, prevention or treatment of cellular proliferation
CC   disorders, cancers, hyperplasias, restenosis, vascular malfunctions,
CC   abnormal wound healing, cardiac hypertrophy, inflammatory and immune
CC   disorders (such as rheumatoid arthritis, gout, psoriasis and diabetic
CC   retinopathy) neurological disorders and disorders of vesicular transport.
CC   This sequence represents the human KinI-3 polypeptide.
XX
SQ   Sequence   1368 AA;

  Query Match             11.2%;  Score 768.5;  DB 23;  Length 1368;
  Best Local Similarity   47.5%;  Pred. No. 5.3e-35;
  Matches  164;  Conservative   58;  Mismatches  116;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| ||||||   | :: : :||||::  || : | :  ||:|:|| :| |  |:|| :
Db        223 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 282

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                 |  ||  |  |||  :: ||   :||||||||:||||||:|:      : ||::  |
Db        283 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 336

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              | |||  | :        ::|||||||||:||||| :|| : | |: |  | :  |: |:
Db        337 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 396

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              | : | |:  :: |   |  |    | :||||||:: | :||  | |  |:|:|||||||
Db        397 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 455

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              || ||    ::||: :|| ||:|||||||||||:| :  | |||:|:||:||:| |:| :
Db        456 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 515

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTN 348
              |: |||||||:    | |||||||: |||  |    |       |
Db        516 KTCMIANISPSHVATEHTLNTLRYADRVKELKKGIKCCTSVTSRN 560


RESULT 10
AAU77185
ID   AAU77185 standard; Protein; 343 AA.
XX
AC   AAU77185;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Human KinI-3 DNA fragment with flanking vector sequences #3.
XX
KW   Human; kinesin motor protein; KinI-3; microtubule; hyperplasia;
KW   depolymerisation; cellular proliferation disorder; cancer; restenosis;
KW   vascular malfunction; abnormal wound healing; cardiac hypertrophy; gout;
KW   inflammatory disorder; immune disorder; rheumatoid arthritis; psoriasis;
KW   diabetic retinopathy; neurological disorder; immunostimulant; vasotropic;
KW   vesicular transport disorder; immunosuppressive; antiinflammatory;
KW   cytostatic.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200226929-A2.
XX
PD   04-APR-2002.
XX
PF   28-SEP-2001; 2001WO-US30750.
XX
PR   29-SEP-2000; 2000US-0675227.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Beraud C,  Guo J,  Freedman R,  Patel UA,  Davies KA;
XX
DR   WPI; 2002-352124/38.
DR   N-PSDB; ABK48000.
XX
PT   An novel isolated microtubule motor protein, KinI-3 used for preparing
PT   compositions for the diagnosis, prevention or treatment of diseases
PT   such as cancers, hyperplasias or restenosis -
XX
PS   Disclosure; Fig 5; 68pp; English.
XX
CC   The invention relates to a human kinesin motor protein (KinI-3), with
CC   microtubule depolymerisation activity, and the polynucleotide encoding
CC   it. The novel KinI-3 motor protein can be used to prepare compositions
CC   for the diagnosis, prevention or treatment of cellular proliferation
CC   disorders, cancers, hyperplasias, restenosis, vascular malfunctions,
CC   abnormal wound healing, cardiac hypertrophy, inflammatory and immune
CC   disorders (such as rheumatoid arthritis, gout, psoriasis and diabetic
CC   retinopathy) neurological disorders and disorders of vesicular transport.
CC   This sequence represents a human KinI-3 polypeptide fragment with
CC   flanking vector sequences.
XX
SQ   Sequence   343 AA;

  Query Match             11.1%;  Score 758.5;  DB 23;  Length 343;
  Best Local Similarity   48.9%;  Pred. No. 4.1e-35;
  Matches  161;  Conservative   58;  Mismatches  103;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| ||||||   | :: : :||||::  || : | :  ||:|:|| :| |  |:|| :
Db         12 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 71

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                 |  ||  |  |||  :: ||   :||||||||:||||||:|:      : ||::  |
Db         72 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 125

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
              | |||  | :        ::|||||||||:||||| :|| : | |: |  | :  |: |:
Db        126 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 185

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              | : | |:  :: |   |  |    | :||||||:: | :||  | |  |:|:|||||||
Db        186 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 244

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              || ||    ::||: :|| ||:|||||||||||:| :  | |||:|:||:||:| |:| :
Db        245 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 304

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVK 332
              |: |||||||:    | |||||||: |||
Db        305 KTCMIANISPSHVATEHTLNTLRYADRVK 333


RESULT 11
AAU77183
ID   AAU77183 standard; Protein; 373 AA.
XX
AC   AAU77183;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Human KinI-3 DNA fragment with flanking vector sequences #1.
XX
KW   Human; kinesin motor protein; KinI-3; microtubule; hyperplasia;
KW   depolymerisation; cellular proliferation disorder; cancer; restenosis;
KW   vascular malfunction; abnormal wound healing; cardiac hypertrophy; gout;
KW   inflammatory disorder; immune disorder; rheumatoid arthritis; psoriasis;
KW   diabetic retinopathy; neurological disorder; immunostimulant; vasotropic;
KW   vesicular transport disorder; immunosuppressive; antiinflammatory;
KW   cytostatic.
XX
OS   Homo sapiens.
```

```
OS   Synthetic.
XX
PN   WO200226929-A2.
XX
PD   04-APR-2002.
XX
PF   28-SEP-2001; 2001WO-US30750.
XX
PR   29-SEP-2000; 2000US-0675227.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Beraud C,  Guo J,  Freedman R,  Patel UA,  Davies KA;
XX
DR   WPI; 2002-352124/38.
DR   N-PSDB; ABK47998.
XX
PT   An novel isolated microtubule motor protein, KinI-3 used for preparing
PT   compositions for the diagnosis, prevention or treatment of diseases
PT   such as cancers, hyperplasias or restenosis -
XX
PS   Disclosure; Fig 3; 68pp; English.
XX
CC   The invention relates to a human kinesin motor protein (KinI-3), with
CC   microtubule depolymerisation activity, and the polynucleotide encoding
CC   it. The novel KinI-3 motor protein can be used to prepare compositions
CC   for the diagnosis, prevention or treatment of cellular proliferation
CC   disorders, cancers, hyperplasias, restenosis, vascular malfunctions,
CC   abnormal wound healing, cardiac hypertrophy, inflammatory and immune
CC   disorders (such as rheumatoid arthritis, gout, psoriasis and diabetic
CC   retinopathy) neurological disorders and disorders of vesicular transport.
CC   This sequence represents a human KinI-3 polypeptide fragment with
CC   flanking vector sequences.
XX
SQ   Sequence   373 AA;

  Query Match             11.1%;  Score 758.5;  DB 23;  Length 373;
  Best Local Similarity   48.9%;  Pred. No. 4.5e-35;
  Matches  161;  Conservative   58;  Mismatches  103;  Indels    7;  Gaps    3;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              ||:| |||||    | :: : :||||::  || : | :  ||:|:|| :| |   |:|| :
Db         42 KIRVCVRKRPLGMREVRRGEINIITVEDKETLLVHEKKEAVDLTQYILQHVFYFDEVFGE 101

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA 123
                 |  ||  |  |||  :: ||   :||||||||:||||||:|:        : ||:: |
Db        102 ACTNQDVYMKTTHPLIQHIF-NGGNATCFAYGQTGAGKTYTMIGTH-----ENPGLYALA 155

Qy        124 AGDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILR 183
               | |||  | :          ::|||||||||:||||| :||  : | |:  |   |  : |:  |:
Db        156 AKDIFRQLEVSQPRKHLFVWISFYEIYCGQLYDLLNRRKRLFAREDSKHMVQIVGLQELQ 215

Qy        184 VLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGS 243
              | : |  |:  ::  |    |    |     | :||||||:: | :||  |   |:|:|||||||
Db        216 VDSVELLLEVILKGSKERSTGATGVNADSSRSHAVIQIQIKDSAKRT-FGRISFIDLAGS 274

Qy        244 ERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKS 303
              || ||     ::||: :|| ||:||||||||||||:| :    ||| |:||:||:| |:| :
Db        275 ERAADARDSDRQTKMEGAEINQSLLALKECIRALDQEHTHTPFRQSKLTQVLKDSFIGNA 334

Qy        304 KSIMIANISPTISCCEQTLNTLRYSSRVK 332
              |: |||||||:     | ||||||: |||
Db        335 KTCMIANISPSHVATEHTLNTLRYADRVK 363

RESULT 12
AAY49949
ID   AAY49949 standard; Protein; 730 AA.
XX
AC   AAY49949;
XX
DT   02-FEB-2000  (first entry)
XX
DE   Xenopus laevis kinesin central motor 1.
XX
KW   Strongylocentrotus purpuratus; katanin; p60 subunit; p80 subunit;
KW   microtubule severing protein; ATPase; detection; identification;
KW   microtubule depolymerisation inhibitor; screening; cell cycle;
KW   kinesin central motor 1; XKCM1; regulation; antimitotic.
XX
OS   Xenopus laevis.
XX
PN   WO9953295-A1.
XX
PD   21-OCT-1999.
XX
PF   13-APR-1999;   99WO-US08086.
XX
PR   14-APR-1998;   98US-0081734.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Vale RD,  Hartman JJ;
XX
DR   WPI; 2000-013108/01.
DR   N-PSDB; AAZ35759.
XX
PT   Identifying agents that modulate microtubule depolymerisation  -
XX
PS   Disclosure; Page 66; 69pp; English.
XX
CC   The present invention describes a method for identifying a modulator of
CC   microtubule (MT) depolymerisation by contacting a polymerised MT with
CC   an MT severing or depolymerising protein and a candidate agent in the
CC   presence of ATP or GTP, and detecting formation of tubulin monomers,
CC   dimmers or oligomers. The method is useful to detect agents that
CC   regulate the cytoskeleton and the cell cycle. Such agents have potential
CC   for use as anti-mitotic agents. The present sequence represents the
CC   Xenopus laevis kinesin central motor 1 (XKCM1), which regulates
CC   microtubule dynamics during mitotic spindle assembly. XKCM1 localises
CC   to centromeres and appears to regulate the polymerisation dynamics of
CC   microtubules, and so is suitable for use in the method of the invention.
XX
SQ   Sequence   730 AA;

  Query Match             11.1%;  Score 758.5;  DB 21;  Length 730;
  Best Local Similarity   42.1%;  Pred. No. 9.6e-35;
  Matches  175;  Conservative   84;  Mismatches  130;  Indels   27;  Gaps   12;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | |||||:: |  ||: |||:|  :   : : ||: |||:|||:|   |   |  ||:
Db        262 RICVCVRKRPLNKQELSKKEIDIISVPSKNIVLVHEPKLKVDLTKYLENQAFRFDFSFDE 321

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |  |  ||  | :||:  ::| |    :|||||||||||:|| |        |: : |:: :
Db        322 TATNEVVYRFTARPLVQSIFEGG-KATCFAYGQTGSGKTHTMGGDFSGKSQNVSKGVYAF 380

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
              |: |:|  |   |:   |      :|::|:||| ||::||| |:  :  ||: |:|| |
Db        381 ASRDVFLLL---DQPRYKHLDLDVFVTFFEIYNGKVFDLLNKKTKLRVLEDAKQEVQVVG 437

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSL-GKIA 236
              |   :|:: :: : |||: |   |     |   ||||| | | |:   : : | ||  :
Db        438 LLEKQVISADD-VFKMIEIGSACRTSGQTFANTSSSRSHACLQIILR---RGSKLHGKFS 493

Qy        237 FIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLR 296
               :||||:||| || |  : :|:  || ||||||||||||||:  :|:| |||:|:|:|::||
Db        494 LVDLAGNERGVDTASADRITRMKGAEINRSLLALKECIRALGQNKSHTPFRESKLTQILR 553

Qy        297 DIFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISIL 355
              | |:|: |:: ||| :|| : || ||||||: |||        ::          :|| : : :
Db        554 DSFIGENSRTCMIAMLSPGFNSCEYTLNTLRYADRVKELSPQNA------ETNDDNLQME 607

Qy        356 DSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
              || ||     :|||  : :: :  || :    :   |     | :  |    |::|: |: ||
Db        608 DSGGSH---ASIEGLQLQDDFLLKDEELSTHN--SFQDALNRVGELEDKAVDELRE 658
```

```
RESULT 13
AAE14504
ID   AAE14504 standard; Protein; 405 AA.
XX
AC   AAE14504;
XX
DT   09-APR-2002  (first entry)
XX
DE   Human mitotic centromere-associated kinesin protein fragment #6.
XX
KW   Human; MCAK; mitotic centromere-associated kinesin; motor domain;
KW   cellular proliferation disorder; cancer; hyperplasia; restenosis;
KW   cardiac hypertrophy; immune disorder; inflammation; autoimmune disease;
KW   arthritis; graft rejection; inflammatory bowel disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Encoded by AC"
XX
PN   US6331424-B1.
XX
PD   18-DEC-2001.
XX
PF   15-JUN-2000; 2000US-0594669.
XX
PR   20-APR-1999;   99US-0295612.
PR   18-MAY-1999;   99US-0314464.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Beraud C,  Sakowicz R;
XX
DR   WPI; 2002-089075/12.
DR   N-PSDB; AAD24085.
XX
PT   New human MCAK (mitotic centromere-associated kinesin) protein useful
PT   in identifying agents for use in the treatment of cellular
PT   proliferation disorders -
XX
PS   Claim 1; Fig 12; 44pp; English.
XX
CC   The invention relates to human MCAK (mitotic centromere-associated
CC   kinesin) protein, and its fragments that comprise a motor domain and
CC   directly or indirectly produce ADP or phosphate. The MCAK enzyme and
CC   its fragments are used in methods to identify compounds that modulate
CC   their activity. Modulators of MCAK are useful as therapeutic agents
CC   for treating cellular proliferation disorders such as cancer,
CC   hyperplasias, restenosis, cardiac hypertrophy, immune disorders,
CC   inflammation, autoimmune disease, arthritis, graft rejection,
CC   inflammatory bowel disease. The present sequence is a human MCAK
CC   fragment.
XX
SQ   Sequence   405 AA;

  Query Match             11.0%;  Score 752.5;  DB 23;  Length 405;
  Best Local Similarity   42.1%;  Pred. No. 1.1e-34;
  Matches  174;  Conservative   75;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                :| | ||||||:: |  ||: |:|:: :  | |  : ||: |||:|||:|    |   | ||:
Db          3 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 62

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |  |  ||  | :||:  ::| |   :|||||||||||:|| |       |: : ||:
Db         63 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 121

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|   |   | |   : ::::|:||| |||:|||  |:  :  ||:||::| |  |:
Db        122 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 180

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                | : :: ::|||| |   |  |     |  |||||   | |:    |     || : :|
Db        181 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 237

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||  | ::||: :|| ||:|||||||||||:  :| | |||:|:||:|||| |
Db        238 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 297

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||       |      |:          :: :
Db        298 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 348

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                  || |  |   |   :   :|              |: : :    |: |: :: :|
Db        349 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 399


RESULT 14
AAE14503
ID   AAE14503 standard; Protein; 434 AA.
XX
AC   AAE14503;
XX
DT   09-APR-2002  (first entry)
XX
DE   Human mitotic centromere-associated kinesin protein fragment #5.
XX
KW   Human; MCAK; mitotic centromere-associated kinesin; motor domain;
KW   cellular proliferation disorder; cancer; hyperplasia; restenosis;
KW   cardiac hypertrophy; immune disorder; inflammation; autoimmune disease;
KW   arthritis; graft rejection; inflammatory bowel disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Encoded by CA"
XX
PN   US6331424-B1.
XX
PD   18-DEC-2001.
XX
PF   15-JUN-2000; 2000US-0594669.
XX
PR   20-APR-1999;   99US-0295612.
PR   18-MAY-1999;   99US-0314464.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Beraud C,  Sakowicz R;
XX
DR   WPI; 2002-089075/12.
DR   N-PSDB; AAD24084.
XX
PT   New human MCAK (mitotic centromere-associated kinesin) protein useful
PT   in identifying agents for use in the treatment of cellular
PT   proliferation disorders -
XX
PS   Claim 1; Fig 10; 44pp; English.
XX
CC   The invention relates to human MCAK (mitotic centromere-associated
CC   kinesin) protein, and its fragments that comprise a motor domain and
CC   directly or indirectly produce ADP or phosphate. The MCAK enzyme and
CC   its fragments are used in methods to identify compounds that modulate
CC   their activity. Modulators of MCAK are useful as therapeutic agents
CC   for treating cellular proliferation disorders such as cancer,
CC   hyperplasias, restenosis, cardiac hypertrophy, immune disorders,
CC   inflammation, autoimmune disease, arthritis, graft rejection,
CC   inflammatory bowel disease. The present sequence is a human MCAK
CC   fragment.
XX
SQ   Sequence   434 AA;
```

```
  Query Match             11.0%;  Score 752.5;  DB 23;  Length 434;
  Best Local Similarity   42.1%;  Pred. No. 1.2e-34;
  Matches  174;  Conservative   75;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|    |   |  ||:
Db         32 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 91

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |  |  ||  | :||:  ::| |   :|||||||||||:|| |        |: : ||:
Db         92 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 150

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |    | |   : ::::|:||| |||:|||  |:  :  ||:||::| |  |:
Db        151 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 209

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                 | : :: ::||||  |      |    |      ||||||   | |:     |     || : :|
Db        210 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 266

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||| | ::||: :|| ||:|||||||||||:    :| | |||:|:||:|||| |
Db        267 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 326

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || ||||||: |||      |     |:          ::  :
Db        327 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 377

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                  || | |   |   :     :|                |: : :    |: |: :: :|
Db        378 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 428
```

RESULT 15
AAE14502
ID AAE14502 standard; Protein; 473 AA.
XX
AC AAE14502;
XX
DT 09-APR-2002 (first entry)
XX
DE Human mitotic centromere-associated kinesin protein fragment #4.
XX
KW Human; MCAK; mitotic centromere-associated kinesin; motor domain;
KW cellular proliferation disorder; cancer; hyperplasia; restenosis;
KW cardiac hypertrophy; immune disorder; inflammation; autoimmune disease;
KW arthritis; graft rejection; inflammatory bowel disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 2
FT /note= "Encoded by GG"
XX
PN US6331424-B1.
XX
PD 18-DEC-2001.
XX
PF 15-JUN-2000; 2000US-0594669.
XX
PR 20-APR-1999; 99US-0295612.
PR 18-MAY-1999; 99US-0314464.
XX
PA (CYTO-) CYTOKINETICS INC.
XX
PI Beraud C, Sakowicz R;
XX
DR WPI; 2002-089075/12.
DR N-PSDB; AAD24083.
XX
PT New human MCAK (mitotic centromere-associated kinesin) protein useful
PT in identifying agents for use in the treatment of cellular
PT proliferation disorders -
XX
PS Claim 1; Fig 8; 44pp; English.
XX
CC The invention relates to human MCAK (mitotic centromere-associated
CC kinesin) protein, and its fragments that comprise a motor domain and
CC directly or indirectly produce ADP or phosphate. The MCAK enzyme and
CC its fragments are used in methods to identify compounds that modulate
CC their activity. Modulators of MCAK are useful as therapeutic agents
CC for treating cellular proliferation disorders such as cancer,
CC hyperplasias, restenosis, cardiac hypertrophy, immune disorders,
CC inflammation, autoimmune disease, arthritis, graft rejection,
CC inflammatory bowel disease. The present sequence is a human MCAK
CC fragment.
XX
SQ Sequence 473 AA;

```
  Query Match             11.0%;  Score 752.5;  DB 23;  Length 473;
  Best Local Similarity   42.1%;  Pred. No. 1.3e-34;
  Matches  174;  Conservative   75;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|    |   |  ||:
Db         71 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 130

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |  |  ||  | :||:  ::| |   :|||||||||||:|| |        |: : ||:
Db        131 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 189

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |    | |   : ::::|:||| |||:|||  |:  :  ||:||::| |  |:
Db        190 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 248

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                 | : :: ::||||  |      |    |      ||||||   | |:     |     || : :|
Db        249 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 305

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||| | ::||: :|| ||:|||||||||||:    :| | |||:|:||:|||| |
Db        306 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 365

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || ||||||: |||      |     |:          ::  :
Db        366 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 416

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                  || | |   |   :     :|                |: : :    |: |: :: :|
Db        417 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 467
```

RESULT 16
AAE14505
ID AAE14505 standard; Protein; 723 AA.
XX
AC AAE14505;
XX
DT 09-APR-2002 (first entry)
XX
DE Human mitotic centromere-associated kinesin protein fragment #7.
XX
KW Human; MCAK; mitotic centromere-associated kinesin; motor domain;
KW cellular proliferation disorder; cancer; hyperplasia; restenosis;
KW cardiac hypertrophy; immune disorder; inflammation; autoimmune disease;
KW arthritis; graft rejection; inflammatory bowel disease.
XX
OS Homo sapiens.
XX
PN US6331424-B1.
XX
PD 18-DEC-2001.
XX
PF 15-JUN-2000; 2000US-0594669.
XX

```
PR    20-APR-1999;   99US-0295612.
PR    18-MAY-1999;   99US-0314464.
XX
PA    (CYTO-) CYTOKINETICS INC.
XX
PI    Beraud C,  Sakowicz R;
XX
DR    WPI; 2002-089075/12.
DR    N-PSDB; AAD24086.
XX
PT    New human MCAK (mitotic centromere-associated kinesin) protein useful
PT    in identifying agents for use in the treatment of cellular
PT    proliferation disorders -
XX
PS    Claim 1; Fig 14; 44pp; English.
XX
CC    The invention relates to human MCAK (mitotic centromere-associated
CC    kinesin) protein, and its fragments that comprise a motor domain and
CC    directly or indirectly produce ADP or phosphate. The MCAK enzyme and
CC    its fragments are used in methods to identify compounds that modulate
CC    their activity. Modulators of MCAK are useful as therapeutic agents
CC    for treating cellular proliferation disorders such as cancer,
CC    hyperplasias, restenosis, cardiac hypertrophy, immune disorders,
CC    inflammation, autoimmune disease, arthritis, graft rejection,
CC    inflammatory bowel disease. The present sequence is a human MCAK
CC    fragment.
XX
SQ    Sequence   723 AA;

  Query Match             11.0%;  Score 752.5;  DB 23;  Length 723;
  Best Local Similarity   42.1%;  Pred. No. 2.1e-34;
  Matches  174;  Conservative   75;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |   ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db        256 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 315

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |   |  ||   | :||:  ::| |    :||||||||||||:|| |        |: : ||:
Db        316 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 374

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |   .| |    : ::::|:||| |||:||| |:  :   ||:||::| |   |:
Db        375 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 433

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                  | : :: ::|||| |     |    |      |   ||||||    |  |:    |      || : :|
Db        434 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 490

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||| | ::||: :|| ||:||||||||||: :| |  |||:|:||:|||| |
Db        491 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 550

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||          |     |:            :: :
Db        551 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 601

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                  || | |  |  :    :|                   |: : :    |: |: :: :|
Db        602 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 652

RESULT 17
AAE14506
ID    AAE14506 standard; Protein; 725 AA.
XX
AC    AAE14506;
XX
DT    09-APR-2002  (first entry)
XX
DE    Human mitotic centromere-associated kinesin protein.
XX
KW    Human; MCAK; mitotic centromere-associated kinesin; motor domain;
KW    cellular proliferation disorder; cancer; hyperplasia; restenosis;
KW    cardiac hypertrophy; immune disorder; inflammation; autoimmune disease;
KW    arthritis; graft rejection; inflammatory bowel disease.
XX
OS    Homo sapiens.
XX
PN    US6331424-B1.
XX
PD    18-DEC-2001.
XX
PF    15-JUN-2000; 2000US-0594669.
XX
PR    20-APR-1999;   99US-0295612.
PR    18-MAY-1999;   99US-0314464.
XX
PA    (CYTO-) CYTOKINETICS INC.
XX
PI    Beraud C,  Sakowicz R;
XX
DR    WPI; 2002-089075/12.
DR    N-PSDB; AAD24087.
XX
PT    New human MCAK (mitotic centromere-associated kinesin) protein useful
PT    in identifying agents for use in the treatment of cellular
PT    proliferation disorders -
XX
PS    Disclosure; Fig 16; 44pp; English.
XX
CC    The invention relates to human MCAK (mitotic centromere-associated
CC    kinesin) protein, and its fragments that comprise a motor domain and
CC    directly or indirectly produce ADP or phosphate. The MCAK enzyme and
CC    its fragments are used in methods to identify compounds that modulate
CC    their activity. Modulators of MCAK are useful as therapeutic agents
CC    for treating cellular proliferation disorders such as cancer,
CC    hyperplasias, restenosis, cardiac hypertrophy, immune disorders,
CC    inflammation, autoimmune disease, arthritis, graft rejection,
CC    inflammatory bowel disease. The present sequence is human MCAK protein.
XX
SQ    Sequence   725 AA;

  Query Match             11.0%;  Score 750.5;  DB 23;  Length 725;
  Best Local Similarity   41.9%;  Pred. No. 2.7e-34;
  Matches  173;  Conservative   76;  Mismatches  143;  Indels   21;  Gaps   10;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |   ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db        258 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 317

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |   |  ||   | :||:  ::| |    :||||||||||||:|| |        |: : ||:
Db        318 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 376

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |   | |    : ::::|:||| |||:||| |:  :   ||:||::| |   |:
Db        377 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 435

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                  | : :: ::|:|  |     |    |      |   ||||||    |  |:    |      || : :|
Db        436 EHLVNSADD-VIKMLDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 492

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||| | ::||: :|| ||:||||||||||: :| |  |||:|:||:|||| |
Db        493 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 552

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||          |     |:            :: :
Db        553 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 603

Qy        359 GSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILKNKSFDKPRE 411
                  || | |  |  :    :|                   |: : :    |: |: :: :|
Db        604 TEEMEACS--NGALIPGNLSKEEEELSSQMSSFNEAMTQIRELEEKAMEELKE 654
```

```
RESULT 18
AAB18301
ID   AAB18301 standard; Protein; 1817 AA.
XX
AC   AAB18301;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:159.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 374-379; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   1817 AA;

  Query Match             10.9%;  Score 748.5;  DB 21;  Length 1817;
  Best Local Similarity   23.6%;  Pred. No. 9.8e-34;
  Matches  356;  Conservative  197;  Mismatches  410;  Indels  543;  Gaps   69;

Qy         79 IIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIF-----QYAAGDIFTFLNI 133
Db        391 IFELFTEGCYLFLLPNEKIKSNNIY-----------TPNIFLKKLKNYITSNDFV---I 435

Qy        134 YDKDNTKGIFISFYEIYCGK----------LYDLLQKRKMVAALENGKKEVVVKDLK--- 180
Db        436 HINDRTLNRYIKFLSYHCSNNIYVHILFNDLYITLHKKIFI-----NNYDMLIKHYKSTA 490

Qy        181 --ILRVLTKEELILKMIDGVLLRK-IGVNSQNDESSRSHAILNIDLK-----------DI 226
Db        491 DHILYLTNKNE-NLNYLNTILLNNYYSLYNKKENNEKRQSLENLKVKIFPSHLNYKQHTT 549

Qy        227 NKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPF 286
Db        550 NKNVN----------------DPNQQHKHDKDD---------SYDNTYEQMKNNKNKIYP 584

Qy        287 RDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDD 346
Db        585 NEYITTHILQN---------------------------NYEQNLYSFQKK-------DD 609

Qy        347 TNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKINDKIERNNILK---- 402
Db        610 TNINNIFDL-HKREQIYEYEKEN---ESSDIFRDSYKRKIKEEK-----KKKNIYKYEDH 660

Qy        403 --NKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYN---DNTINKKHNNN 457
Db        661 PLNKEKKKKKFFYINYEKGDDKNDNDL-----------YYNNIYSKNLENIQNKNYNNN 709

Qy        458 NNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNN 517
Db        710 NNNNNNNNNNNNNNNNNNNN-------------NNNNNNNNNNNNNNNNNNNNNFDKY 756

Qy        518 HLPQPNYAFTDTSDFSSL-------DDMNCHLNNNDKSIF-LHKKNLRDNIK-----LKN 564
Db        757 NIPCTNTNLSLLYDKEKLFLFTYAYDKIQTYTYEELKSKYKISTKIVDRNIKMFLKFLKN 816

Qy        565 RSS-----CDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNT------FFKS----- 608
Db        817 YNNNENTYVDNIIS-KKNIFHLLA-SMKNKVT--------NKTNTHKDIYQFIHSWYHIK 866

Qy        609 --NINK----MEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDEN------HND 656
Db        867 LADQNKEHSFQDDKYLINNLYEKHKVQH---------NTMTHHIIKMEDKKGDIHLMENN 917

Qy        657 NKINNGVINIINNSNVNS--------------------INNS----NMNSINNS-NMNS 690
Db        918 NMLLNNNMSLNNNMSLNNSIPLNNSIPLNNSIPLNNSIPLNNSIPLNNSIPLNNSISLNS 977

Qy        691 ---------------------NSIYKSNYNSNQSISDVQIR-YVNEMDTS--------- 718
Db        978 CISLYNSISLYSNKNTSFNHLYNNIYDTCFIQNNYISNQQVQNYKNEKNTNMEHYNEKKL 1037

Qy        719 ----------------------NKNNDNIFF-------------------------- 727
Db       1038 FIYPIYYLEDKNYFLNVVNNIFFNKNYNNTFFYTCQINILSKGLYYFINYYTLLISSNYK 1097

Qy        728 -DAISCDNNMYPNITNNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQP 786
Db       1098 AEEIKTDDNKC-NINNNNNNNNNNNNNNNNN---NNNNNNNNYNNN----NYNNNNVYPL 1149

Qy        787 DNNKNTS----------NIQNINTNK--------------NNQDGNVNYSMNFCH---- 817
Db       1150 INHFTTTFYEMVTYLLKNIYRIHISKFFYIFVALSKFFLMNSYQQSNTNKRENSIHMENV 1209

Qy        818 ------------------YNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQN-----RNDF 853
Db       1210 LYILYIIRKKQYEHVKSILYDKSNENYFRFNENKDIKMENTNMLYNIILNNFSTEDHDEF 1269

Qy        854 EKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKD 913
Db       1270 MTLQKNNEDNNKMIIDNINNVDNIN---------------DLIKSHHCDNNKKED----- 1309

Qy        914 NHIKSNNNNSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDD 973
Db       1310 ------------TSSLHNKLYNGLHFLIMFLNNYLDNTKHFKI----NHFLSSLFYINK- 1352

Qy        974 IIYTINSLNDYMSNTLLH-----FKEKYTYPTLSTNEDIYN-----KEMEGKHIRLDDQD 1023
Db       1353 -IIPPNMKHMYHLETYLHKNHKIYKNKFFY--------IYNGLDLLKKSYLVHIK----- 1398
```

```
Qy       1024 KYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNN--DKNNVDNNNVDNDDDDVDFHNIKN 1081
              |   |       || |        ||||: :| ||    | |:          |::|: ||
Db       1399 KLYINSYIKSYNNKKK-------NNNVNGDVYNNFMYKYNI--------YDNIDYIFIKK 1443

Qy       1082 -----FNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAE----KD 1132
                   : |: | ||      | | : ||      ::| :     ::     |    |    |:
Db       1444 KNLFCYTNHLSLLYF-----TYIYS-LNKFYYCTLYYN----ISKCFYYKINIENIHFKN 1493

Qy       1133 NVIKKYINEDIKNMSLEEIDKTAQSIY---EKRKV------LLTKLLLLFKKNVDTQINN 1183
               :|  :|     | : ::        |||:   | :||       :|: ::|| ||    ::|
Db       1494 KIILFFIFTQCKYIYIKFFRLLVQSIFSSSEFQKVGKLSLYILSNIILLLVKNSRMKLNI 1553

Qy       1184 ETSDLRKDLVMCHICNNNPDDQFHFYAYSRLEKDINLIMLRQIWCE----SENLRLLYQ 1239
               :   : |: :  || :||                  : ||     :::|      |:|| :  :
Db       1554 KKKKIIKN-ISKHIYSNN---------------EFINNNKIKKIHTNNNSMSKNLFICNK 1597

Qy       1240 FLVVEYQ--------------------NKSANSVLLNVSSNNGDIILLNKKLVQDNIKN 1278
                 | :::                    ||     ::| | ::|| :              :|  |
Db       1598 LLNIQWNYIFPMDLFISSNLSHETELIINKLEQNILNNNNNNNN---------NNNNN 1647

Qy       1279 SMDHNN 1284
              :  |||
Db       1648 NNKHNN 1653
```

RESULT 19
AAE14501
ID   AAE14501 standard; Protein; 362 AA.
XX
AC   AAE14501;
XX
DT   09-APR-2002 (first entry)
XX
DE   Human mitotic centromere-associated kinesin protein fragment #3.
XX
KW   Human; MCAK; mitotic centromere-associated kinesin; motor domain;
KW   cellular proliferation disorder; cancer; hyperplasia; restenosis;
KW   cardiac hypertrophy; immune disorder; inflammation; autoimmune disease;
KW   arthritis; graft rejection; inflammatory bowel disease.
XX
OS   Homo sapiens.
XX
FH   Key Location/Qualifiers
FT   Misc-difference 2
FT   /note= "Encoded by AC"
XX
PN   US6331424-B1.
XX
PD   18-DEC-2001.
XX
PF   15-JUN-2000; 2000US-0594669.
XX
PR   20-APR-1999; 99US-0295612.
PR   18-MAY-1999; 99US-0314464.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Beraud C, Sakowicz R;
XX
DR   WPI; 2002-089075/12.
DR   N-PSDB; AAD24082.
XX
PT   New human MCAK (mitotic centromere-associated kinesin) protein useful
PT   in identifying agents for use in the treatment of cellular
PT   proliferation disorders -
XX
PS   Claim 1; Fig 6; 44pp; English.
XX
CC   The invention relates to human MCAK (mitotic centromere-associated
CC   kinesin) protein, and its fragments that comprise a motor domain and
CC   directly or indirectly produce ADP or phosphate. The MCAK enzyme and
CC   its fragments are used in methods to identify compounds that modulate
CC   their activity. Modulators of MCAK are useful as therapeutic agents
CC   for treating cellular proliferation disorders such as cancer,
CC   hyperplasias, restenosis, cardiac hypertrophy, immune disorders,
CC   inflammation, autoimmune disease, arthritis, graft rejection,
CC   inflammatory bowel disease. The present sequence is a human MCAK
CC   fragment.
XX
SQ   Sequence 362 AA;

```
  Query Match             10.9%;  Score 745.5;  DB 23;  Length 362;
  Best Local Similarity   45.7%;  Pred. No. 2.3e-34;
  Matches  168;  Conservative   65;  Mismatches  116;  Indels   19;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |   ||: |:|:: :  | |  : ||: |||:|||:|    |   |  ||:
Db          3 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 62

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
                |  |  ||  | :||:  ::| |   :|||||||||||:|| |         |: : ||:
Db         63 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 121

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |    | |    : ::::|:||| |||:||| |:  :   ||:||::| |  |:
Db        122 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 180

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                 | : :: ::|||| |    |   |     |  ||||||   | |:     |     || : :|
Db        181 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 237

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
                |||:||||||  | ::||: :|| ||:|||||||||||:   :|  |  |||:|:||:|||| |
Db        238 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 297

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||        |       |:              :: :
Db        298 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 348

Qy        359 GSEMNASS 366
                 || | |
Db        349 TEEMEACS 356
```

RESULT 20
AAE14500
ID   AAE14500 standard; Protein; 391 AA.
XX
AC   AAE14500;
XX
DT   09-APR-2002 (first entry)
XX
DE   Human mitotic centromere-associated kinesin protein fragment #2.
XX
KW   Human; MCAK; mitotic centromere-associated kinesin; motor domain;
KW   cellular proliferation disorder; cancer; hyperplasia; restenosis;
KW   cardiac hypertrophy; immune disorder; inflammation; autoimmune disease;
KW   arthritis; graft rejection; inflammatory bowel disease.
XX
OS   Homo sapiens.
XX
FH   Key Location/Qualifiers
FT   Misc-difference 2
FT   /note= "Encoded by CA"
XX
PN   US6331424-B1.
XX
PD   18-DEC-2001.
XX
PF   15-JUN-2000; 2000US-0594669.
XX
PR   20-APR-1999; 99US-0295612.
PR   18-MAY-1999; 99US-0314464.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX

```
PI   Beraud C,  Sakowicz R;
XX
DR   WPI; 2002-089075/12.
DR   N-PSDB; AAD24081.
XX
PT   New human MCAK (mitotic centromere-associated kinesin) protein useful
PT   in identifying agents for use in the treatment of cellular
PT   proliferation disorders -
XX
PS   Claim 1; Fig 4; 44pp; English.
XX
CC   The invention relates to human MCAK (mitotic centromere-associated
CC   kinesin) protein, and its fragments that comprise a motor domain and
CC   directly or indirectly produce ADP or phosphate. The MCAK enzyme and
CC   its fragments are used in methods to identify compounds that modulate
CC   their activity. Modulators of MCAK are useful as therapeutic agents
CC   for treating cellular proliferation disorders such as cancer,
CC   hyperplasias, restenosis, cardiac hypertrophy, immune disorders,
CC   inflammation, autoimmune disease, arthritis, graft rejection,
CC   inflammatory bowel disease. The present sequence is a human MCAK
CC   fragment.
XX
SQ   Sequence   391 AA;

  Query Match             10.9%;  Score 745.5;  DB 23;  Length 391;
  Best Local Similarity   45.7%;  Pred. No. 2.6e-34;
  Matches  168;  Conservative   65;  Mismatches  116;  Indels   19;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db         32 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 91

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |   |  ||   |  :||:   ::|  |    :|||||||||||:||  |        |: : ||:
Db         92 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 150

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |   | |    : ::::|:||| |||:||| |:   :  ||:||::| |   |:
Db        151 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 209

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                | : :: ::|||| |    |    |    |     ||||||    | |:    |      || : :|
Db        210 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 266

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||  | ::||: :|| ||:|||||||||||:   :| | |||:|:||:|||| |
Db        267 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 326

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||      |     |:             ::  :
Db        327 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 377

Qy        359 GSEMNASS 366
                 || | |
Db        378 TEEMEACS 385


RESULT 21
AAE14499
ID   AAE14499 standard; Protein; 430 AA.
XX
AC   AAE14499;
XX
DT   09-APR-2002  (first entry)
XX
DE   Human mitotic centromere-associated kinesin protein fragment #1.
XX
KW   Human; MCAK; mitotic centromere-associated kinesin; motor domain;
KW   cellular proliferation disorder; cancer; hyperplasia; restenosis;
KW   cardiac hypertrophy; immune disorder; inflammation; autoimmune disease;
KW   arthritis; graft rejection; inflammatory bowel disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 2
FT                   /note= "Encoded by GG"
XX
PN   US6331424-B1.
XX
PD   18-DEC-2001.
XX
PF   15-JUN-2000; 2000US-0594669.
XX
PR   20-APR-1999;   99US-0295612.
PR   18-MAY-1999;   99US-0314464.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Beraud C,  Sakowicz R;
XX
DR   WPI; 2002-089075/12.
DR   N-PSDB; AAD24080.
XX
PT   New human MCAK (mitotic centromere-associated kinesin) protein useful
PT   in identifying agents for use in the treatment of cellular
PT   proliferation disorders -
XX
PS   Claim 1; Fig 2; 44pp; English.
XX
CC   The invention relates to human MCAK (mitotic centromere-associated
CC   kinesin) protein, and its fragments that comprise a motor domain and
CC   directly or indirectly produce ADP or phosphate. The MCAK enzyme and
CC   its fragments are used in methods to identify compounds that modulate
CC   their activity. Modulators of MCAK are useful as therapeutic agents
CC   for treating cellular proliferation disorders such as cancer,
CC   hyperplasias, restenosis, cardiac hypertrophy, immune disorders,
CC   inflammation, autoimmune disease, arthritis, graft rejection,
CC   inflammatory bowel disease. The present sequence is a human MCAK
CC   fragment.
XX
SQ   Sequence   430 AA;

  Query Match             10.9%;  Score 745.5;  DB 23;  Length 430;
  Best Local Similarity   45.7%;  Pred. No. 2.8e-34;
  Matches  168;  Conservative   65;  Mismatches  116;  Indels   19;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db         71 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 130

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              |   |  ||   |  :||:   ::|  |    :|||||||||||:||  |        |: : ||:
Db        131 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 189

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
              |: |:|    |   | |    : ::::|:||| |||:||| |:   :  ||:||::| |   |:
Db        190 ASRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 248

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                | : :: ::|||| |    |    |    |     ||||||    | |:    |      || : :|
Db        249 EHLVNSADD-VIKMIDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 305

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:||||||  | ::||: :|| ||:|||||||||||:   :| | |||:|:||:|||| |
Db        306 LAGNERGADTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 365

Qy        300 VGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDSK 358
              :|: |:: ||| ||| || || |||||||: |||      |     |:             ::  :
Db        366 IGENSRTCMIATISPGISSCEYTLNTLRYADRVKELSPHSGPSGEQ---------LIQME 416

Qy        359 GSEMNASS 366
                 || | |
Db        417 TEEMEACS 424
```

```
RESULT 22
ABB59500
ID   ABB59500 standard; Protein; 729 AA.
XX
AC   ABB59500;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 5292.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03603.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 5292; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   729 AA;

  Query Match             10.9%;  Score 745;  DB 22;  Length 729;
  Best Local Similarity   38.6%;  Pred. No. 5.5e-34;
  Matches  193;  Conservative   86;  Mismatches  161;  Indels   60;  Gaps   15;

Qy           4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
               :| | |||||:|  |:  |: |||||  : :| : | | |||:||::| |:|  |  ||:
Db         233 QITVCVRKRPMSRKEENSKNLDIITVPSADSLIVHELRLKVDLTKFLEHHKFRFDYTFDE 292

Qy          64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYG--QSDTPGIFQ 121
                  |  ||::| :|||  ::| |   :|||||||||||||:|| | : :|  |    ||: 
Db         293 ECSNALVYDHTARPLIRTMFEGG-NATCFAYGQTGSGKTHTM-GGEFFGKVQDCGTGIYA 350

Qy         122 YAAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRK-MVAALENGKKEVVVKD 178
                || |:|  ::   | :   | |  ||:|||  |::|||   | |:  ||: :::|||  
Db         351 MAARDVFEEVSRPEYRQMGAK-ITCSFFEIYGTKVFDLLLPNKPMLRVLEDARQQVVVVG 409

Qy         179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237
               |  : | || | :|::|: |   |  |  | |  ||||||:  | |    :    || :|
Db         410 LTEMPV-TKVEDVLRLIEHGSKERTSGQTSANAKSSRSHAVFQIALHFPDSWGPHGKCSF 468

Qy         238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297
               :||||:|||||| | ::||: :|| ||:||||||||||:    :|:||| |:||:||||
Db         469 VDLAGNERGADTQSADRQTRIEGAEINKSLLALKECIRALSRQSSHLPFRGSKLTQVLRD 528

Qy         298 IFVG--KSKSIMIANISPTISCCEQTLNTLRYSSRVKNF--------------KNKSTCI 341
                |||  |:|: ||| |||::|| | ||||||||: |||                ||  :
Db         529 SFVGGKKNKTCMIAMISPSMSCVENTLNTLRYADRVKELIAKEDEHLQSVEGDGEKSPDL 588

Qy         342 NEE---------------DDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKI 386
               |||               :| : : ::|   : |  |  |  :   :  || |  |     
Db         589 NEESEPEMMADEEGDEEPEDEDNQHLTISSEEASSYNNMSYD---MSFNHTL--NILGPS 643

Qy         387 NRGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTL-- 444
                   | :  |:: :|         | :   | :  :  :  :|::  :| :  |:   : : 
Db         644 RNVDIQEVAEQHALLVEN-----LETYAHNFRQLKTDKEIEQYTQNSESALMKLLAMVNR 698

Qy         445 -------YNDNTINKKHNNN 457
                      ||   : |: | |
Db         699 TRDVTHNYNTQKLLKEENQN 718
```

```
RESULT 23
AAM39258
ID   AAM39258 standard; Protein; 728 AA.
XX
AC   AAM39258;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2403.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI58414.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 4; SEQ ID NO 2403; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
```

```
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   728 AA;

  Query Match             10.7%;  Score 733;  DB 22;  Length 728;
  Best Local Similarity   47.9%;  Pred. No. 2.6e-33;
  Matches  162;  Conservative   61;  Mismatches  101;  Indels   14;  Gaps    8;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db        207 RICVCVRKRPLNKKETQMKDLDVITIPSKDVVMVHEPKQKVDLTRYLENQTFRFDYAFDD 266

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
Db        267 SAPNEMVYRFTARPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 325

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
Db        326 AARDVFLMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 382

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237
Db        383 LQEREVKCVED-VLKLIDIGNSCRTSGQTSANAHSSRSHAVFQIILR--RKGKLHGKFSL 439

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297
Db        440 IDLAGNERGADTSSADRQTRLEGAEINKSLLALKECIRALGRNKPHTPFRASKLTQVLRD 499

Qy        298 IFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNF 334
Db        500 SFIGENSRTCMIATISPGMASCENTLNTLRYANRVKEF 537


RESULT 24
AAW88456
ID   AAW88456 standard; Protein; 679 AA.
XX
AC   AAW88456;
XX
DT   10-MAY-1999  (first entry)
XX
DE   Human kinesin-related protein KINRELP.
XX
KW   Kinesin-related protein; KINRELP; human; cancer; apoptosis;
KW   cell proliferation; inflammation; therapy; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    286..293
FT                   /note= "ATP-binding P-loop"
FT   Domain          424..435
FT                   /note= "kinesin motor domain signature"
FT   Modified-site   9..11
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   106..109
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   244..247
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   455..458
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   127..130
FT                   /note= "potential cAMP- and cAMP-dependent protein
FT                          kinase phosphorylation site"
FT   Modified-site   208..211
FT                   /note= "potential cAMP- and cAMP-dependent protein
FT                          kinase phosphorylation site"
FT   Modified-site   4..7
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   27..30
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   108..111
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   187..190
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   273..280
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   338..341
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   427..430
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   441..444
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   447..450
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   582..585
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   634..637
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   233..241
FT                   /note= "potential tyrosine kinase phosphorylation
FT                          site"
FT   Modified-site   329..336
FT                   /note= "potential tyrosine kinase phosphorylation
FT                          site"
FT   Modified-site   58..60
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   73..75
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   88..90
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   126..128
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   246..248
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   266..268
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   290..292
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   301..303
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   344..346
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   392..394
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   405..407
```

```
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   516..518
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   608..610
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   656..658
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
XX
PN   WO9858064-A1.
XX
PD   23-DEC-1998.
XX
PF   19-JUN-1998;   98WO-US12856.
XX
PR   19-JUN-1997;   97US-0878865.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Au-Young J,  Lal P,  Shah P;
XX
DR   WPI; 1999-080955/07.
DR   N-PSDB; AAX06945.
XX
PT   New kinesin-related protein - useful for treating cancer and
PT   inflammation
XX
PS   Claim 1; Page 49-50; 72pp; English.
XX
CC   This is the amino acid sequence of human kinesin-related protein
CC   (KINRELP), as deduced from a consensus cDNA sequence (see AAX06945).
CC   KINRELP shares 97% identity with murine kinesin-related protein
CC   KIF2.  The invention provides expression vectors, host cells,
CC   agonists, antibodies and antagonists, as well as methods for
CC   treating disorders associated with expression of KINRELP.  KINRELP
CC   and its agonists are used to stimulate cell proliferation, and to
CC   treat a disorder associated with increased apoptosis.  Antagonists
CC   are used to treat cancer and inflammation (claimed).  Particularly,
CC   KINRELP or its fragment may be used to prevent or treat: AIDS and
CC   other infectious or genetic immunodeficiencies, neurodegenerative
CC   diseases such as Alzheimer's disease, Parkinson's disease,
CC   amyotrophic lateral sclerosis, retinitis pigmentosis and
CC   cerebellar degeneration; myelodysplastic syndromes such as aplastic
CC   anaemia; ischemic diseases such as myocardial infarction, stroke
CC   and reperfusion injury; toxin-induced diseases such as
CC   alcohol-induced liver damage, cirrhosis and lathyrism; wasting
CC   diseases such as cachexia and osteoporosis; viral infections such
CC   as those caused by hepatitis B and C; and diseases associated with
CC   inflammation including adult respiratory distress syndrome,
CC   allergy, asthma, arteriosclerosis, bronchitis, emphysema,
CC   hypereosinophilia, myocardial or pericardial inflammation,
CC   rheumatoid arthritis, Addison's disease, AIDS, anaemia,
CC   atherosclerosis, diseases of the digestive system, atopic
CC   dermatitis, dermatomyositis, diabetes mellitus, glomerulonephritis,
CC   gout, Grave's disease, lupus erythematosus, multiple sclerosis,
CC   myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis,
CC   polycystic kidney disease, polymyositis, scleroderma, Sjorgren's
CC   syndrome, autoimmune thyroiditis, complications of cancer,
CC   extracorporeal circulation, viral, bacterial, fungal, parasitic,
CC   protozoal and helminthic infections and trauma.
XX
SQ   Sequence   679 AA;

  Query Match             10.7%;  Score 732.5;  DB 20;  Length 679;
  Best Local Similarity   45.5%;  Pred. No. 2.6e-33;
  Matches  166;  Conservative   63;  Mismatches  111;  Indels   25;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                :| | ||||||:: | : || |:||: :    : : ||: |||:|:|:|   |   | |||
Db        196 RICVCVRKRPLNKKETQMKDLDVITIPSKDVVMVHEPKQKVDLTRYLENQTFRFDYAFDD 255

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              :  |  ||   | :||:  ::| | : :|||||||||||:|| |           |   : ||:
Db        256 SAPNEMVYRFTARPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 314

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
              || |:|  |       | | |         :: :|:||| ||::||| ::   :   ||:||::| |
Db        315 AARDVFLMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 371

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237
              |:    |     |: :||:|| |    |   |   |  |   ||||||:  | |:    |      || :
Db        372 LQEREVKCVED-VLKLIDIGNSCRTSGQTSANAHSSRSHAVFQIILR--RKGKLHGKFSL 428

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297
              |||||:||||||  | ::||: :|| ||:||||||||||||:    :|  | ||| |:||:||||
Db        429 IDLAGNERGADTSSADRQTRLEGAEINKSLLALKECIRALGRNKPHTPFRASKLTQVLRD 488

Qy        298 IFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCI-----------NEED 345
               |:|: |::  |||  |||  ::  || |||||||::|||          |              |: |
Db        489 SFIGENSRTCMIATISPGMASCENTLNTLRYANRVKELTVDPTAAGDVRPIMHHPPNQID 548

Qy        346 DTNTE 350
              |  |:
Db        549 DLETQ 553
```

```
RESULT 25
AAM39257
ID   AAM39257 standard; Protein; 690 AA.
XX
AC   AAM39257;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2402.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI58413.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 4; SEQ ID NO 2402; 10078pp; English.
```

```
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   690 AA;

  Query Match             10.7%;  Score 732.5;  DB 22;  Length 690;
  Best Local Similarity   45.5%;  Pred. No. 2.6e-33;
  Matches  166;  Conservative   63;  Mismatches  111;  Indels   25;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | |||||:: | : || |:||: :    : : ||: |||:|:|:|    |   |  |||
Db        207 RICVCVRKRPLNKKETQMKDLDVITIPSKDVVMVHEPKQKVDLTRYLENQTFRFDYAFDD 266

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              :  |  ||   | :||:  ::| | : :||||||||||||:|| |         |  : ||:
Db        267 SAPNEMVYRFTARPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 325

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
              || |:|  |      | | |        :: :|:||| ||::||| ::   :   ||:||::| |
Db        326 AARDVFLMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 382

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237
              |:    |    |: :|:|| |    |   |   |  | |||||:  | |:    |     || :
Db        383 LQEREVKCVED-VLKLIDIGNSCRTSGQTSANAHSSRSHAVFQIILR--RKGKLHGKFSL 439

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297
              |||||:|||||| | ::|| :|| ||:||||||||||:   :| | ||| |:||:||||
Db        440 IDLAGNERGADTSSADRQTRLEGAEINKSLLALKECIRALGRNKPHTPFRASKLTQVLRD 499

Qy        298 IFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCI-----------NEED 345
               |:|: |:: ||| |:|| :: || |||||||::|||       |            |: |
Db        500 SFIGENSRTCMIATISPGMASCENTLNTLRYANRVKELTVDPTAAGDVRPIMHHPPNQID 559

Qy        346 DTNTE 350
               |  |:
Db        560 DLETQ 564

RESULT 26
AAM41043
ID   AAM41043 standard; Protein; 694 AA.
XX
AC   AAM41043;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 5974.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI60199.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 5974; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   694 AA;

  Query Match             10.7%;  Score 732.5;  DB 22;  Length 694;
  Best Local Similarity   45.5%;  Pred. No. 2.6e-33;
  Matches  166;  Conservative   63;  Mismatches  111;  Indels   25;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | |||||:: | : || |:||: :    : : ||: |||:|:|:|    |   |  |||
Db        211 RICVCVRKRPLNKKETQMKDLDVITIPSKDVVMVHEPKQKVDLTRYLENQTFRFDYAFDD 270

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              :  |  ||   | :||:  ::| | : :||||||||||||:|| |         |  : ||:
Db        271 SAPNEMVYRFTARPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 329

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
              || |:|  |      | | |        :: :|:||| ||::||| ::   :   ||:||::| |
Db        330 AARDVFLMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 386

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237
              |:    |    |: :|:|| |    |   |   |  | |||||:  | |:    |     || :
Db        387 LQEREVKCVED-VLKLIDIGNSCRTSGQTSANAHSSRSHAVFQIILR--RKGKLHGKFSL 443

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297
              |||||:|||||| | ::|| :|| ||:||||||||||:   :| | ||| |:||:||||
Db        444 IDLAGNERGADTSSADRQTRLEGAEINKSLLALKECIRALGRNKPHTPFRASKLTQVLRD 503

Qy        298 IFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCI-----------NEED 345
               |:|: |:: ||| |:|| :: || |||||||::|||       |            |: |
```

```
Db        504 SFIGENSRTCMIATISPGMASCENTLNTLRYANRVKELTVDPTAAGDVRPIMHHPPNQID 563

Qy        346 DTNTE 350
              |  |:
Db        564 DLETQ 568


RESULT 27
AAM41044
ID   AAM41044 standard; Protein; 694 AA.
XX
AC   AAM41044;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 5975.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI60200.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 5975; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   694 AA;

  Query Match             10.7%;  Score 732.5;  DB 22;  Length 694;
  Best Local Similarity   45.5%;  Pred. No. 2.6e-33;
  Matches  166;  Conservative   63;  Mismatches  111;  Indels   25;  Gaps    9;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: | : || |:||: :    : : ||: |||:|:|:|   |  |  |||
Db        211 RICVCVRKRPLNKKETQMKDLDVITIPSKDVVMVHEPKQKVDLTRYLENQTFRFDYAFDD 270

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
              :  |  ||  | :||:  ::| | : :||||||||||||:|| |        |  : ||:
Db        271 SAPNEMVYRFTARPLVETIFERG-MATCFAYGQTGSGKTHTMGGDFSGKNQDCSKGIYAL 329

Qy        123 AAGDIFTFLNIYDKDNTK----GIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKD 178
              || |:|  |     | | |    :: :|:||| ||::||| ::  :   ||:||::| |
Db        330 AARDVFLMLK---KPNYKKLELQVYATFFEIYSGKVFDLLNRKTKLRVLEDGKQQVQVVG 386

Qy        179 LKILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAF 237
              |:   |    |: :||:|| |    |   |   |  |  ||||||:  | |:   |    || :
Db        387 LQEREVKCVED-VLKLIDIGNSCRTSGQTSANAHSSRSHAVFQIILR--RKGKLHGKFSL 443

Qy        238 IDLAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRD 297
              |||||:||||||  | ::||: :|| ||:||||||||||:   :| |  ||| |:||:||||
Db        444 IDLAGNERGADTSSADRQTRLEGAEINKSLLALKECIRALGRNKPHTPFRASKLTQVLRD 503

Qy        298 IFVGK-SKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCI-----------NEED 345
               |:|: |:: ||| ||| :: || ||||||::|||           |               |: |
Db        504 SFIGENSRTCMIATISPGMASCENTLNTLRYANRVKELTVDPTAAGDVRPIMHHPPNQID 563

Qy        346 DTNTE 350
              |  |:
Db        564 DLETQ 568


RESULT 28
AAB18199
ID   AAB18199 standard; Protein; 1436 AA.
XX
AC   AAB18199;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:56.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 133-137; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
```

```
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   1436 AA;

  Query Match             10.2%;  Score 698;  DB 21;  Length 1436;
  Best Local Similarity   23.8%;  Pred. No. 5.3e-31;
  Matches  354;  Conservative  224;  Mismatches  524;  Indels  388;  Gaps  71;

Qy     17 LEKKKKD-----SDIITVKNNCTLYIDEPRYKVDMTKYIERHEFI--VDK----VFDDTV 65
Db     47 IDDKKKDLGKSKYDIFTTDSLSTT-TDEVSYSYQIENKNEEKEYLRYYDKQGGIIRQDNN 105

Qy     66 DNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYA-A 124
Db    106 NN----ENNNNNICNNDHNNNNICN----------NENMLTTKKNDNTIINSNIKYLNN 150

Qy    125 GDIFTFLNIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALE-----NGKKEV--VVK 177
Db    151 NNIFNTNMVPQKNHTQ-----IFNPYDKSMRNIQLYNKAVSFLKNDGDINSKKNTHDNLM 205

Qy    178 DLKILRVLTKEELIL---------KMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDIN 227
Db    206 FLKNIRSKSNNNLIVNRKITNHVTNNVISGMTNKVIGGMASGMTNNVTSSITNNMTSSMT 265

Qy    228 KNTSLGKIAFIDLAGSERGADTVSQN---KQTQTDGANINRSLLALKECIRAMDSDKNHI 284
Db    266 NNMASGMTSSMTNNMASGMTSSITNNMTSSMTNNMASGMTSSI---------------- 308

Qy    285 PFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEE 344
Db    309 ---TNNMTSSMTNNMASSMTSSMTNNMTSSMT------NNMTSSMTNNMLNNMNRVVTNN 359

Qy    345 DDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGK----INDK--IERN 398
Db    360 IITNMNR-SVSGSKSINMSNLLIINKMDYGNDIYHNNNNNNNNSSSGSNIVSGKYFVNSQ 418

Qy    399 NILKNKSFDKPREGFTSTFGKYSSLNDIDK-------------IKKNKKKGLINYKSTLY 445
Db    419 NSSKNNFFTKVGE---STIRSPTNILDIYKQGNMYMHIPKNADLMNNVSSYSIAHENYIK 475

Qy    446 NDNTINKKH--NNN-----NNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNN 498
Db    476 RDNT-NVTHVLNNNHLVNINNVVNNNNLNNNNNLNNNNNLNS-----------NNNLNSN 523

Qy    499 INVNNNNNNNNNNNNSHNNHL---PQPNYAFTDTSDFS---SLDDMNCHLNNNDKSIFLH 552
Db    524 NNLNNNNNLINNNNLINNNYVRNNQAVNNAHTLNAHFNKSDNVDNMRNHIPNND------ 577

Qy    553 KKNLRDNIKLKNRSSCDN---IMNKKKNNLHLARHSVGSKLTMFSYDPQKNKD-NTFFK- 607
Db    578 NKNIVNMLNLKNMKSINDLSVLINKNKPIHHVINGTEVQQKRSLS-NVQKLKTLNTFPNA 636

Qy    608 ----SNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGV 663
Db    637 KGRFSLINKMA--------SMPNMSTTSSMNMSGLNTSSSEGLTNIINMNNINSVNN-- 685

Qy    664 INIINN----SNVNSINN----SNMNSINNSNMNSNSIYKSNYNSNQSISDV-QIRYVNE 714
Db    686 INSVNNINSVNNINSVNNLNSVNNINSVNNINSVNNINSVNNINSVNNINSVNNINSVNN 745

Qy    715 MDTSNKNNDNIFFDAISCDNNM-YPNITNNNNNNNNNN-------NNNNNI-------- 758
Db    746 LNSVNNINSVNNINNINYINNINYVNMNKGLNPINNVSNISSLKLLNNNNDIKKKFNTYG 805

Qy    759 ------------------DVENYNNRDGTNNSMKLYAYNS----------HNLF--QP 786
Db    806 KSEASENLSKNVKYIKYIQENIKYLNNLDDNKRKYSLTSINDVGCIKKKKNMNDLFLGKH 865

Qy    787 DNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNI 846
Db    866 DNMLRTDEIPKINLGKNILNNN-----KIINYNDNDKSNII---------------NNV 904

Qy    847 IQNR------NDFE-KKKKTNFYN----NNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDN 895
Db    905 INKNISTDLVNDREGDMNKMNIHNREKDENNYI----NIGDNKIKK---------NQIDV 951

Qy    896 MKNNEMKNNEMKDNEM-----------KDNHIKSNNNNSSSSSSSNN----NIYNNINDD 940
Db    952 VNNKVMKLDNMEDEEAMNKLSLISLYPNNNHIINNVNNVNNVNNVNNVNNVNNVNNVNNV 1011

Qy    941 DTFQN-DYCHN-DNTFTIRRKNNTNINSNIYQNDDI--IYTINSLNDYMSNTLLHFKEKY 996
Db   1012 NNVNNVNYMNNMNNVNNVNNMNNVNNMNNVNNVNSVNNIKGINNMNNNNNNINMNRSYKM 1071

Qy    997 TYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKN---------------- 1039
Db   1072 NMKKVSKKDNGQNVVSEKRF----SEEKY--NFLKNLIRNNKNMVKLKYLNKFLGKRSGP 1125

Qy   1040 NVDNNVD------NNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQK 1093
Db   1126 SIKNNMNDMMVKMNNNMKDIMHIKDATNI--NKINN--------KLVNLNTNNCISYNSC 1175

Qy   1094 NVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSLEEIDK 1153
Db   1176 NKMNYIHKCKKKRVLCL---DTKHGKNEI--------KQNEKLIYTNYEIKMFLLNTI-K 1223

Qy   1154 TAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHI-CNNNPDDQFHFYAYS 1212
Db   1224 AIGIVFKKWKFKNFGLYFWY---------------------HIKCIENERD-LNFYI-- 1258

Qy   1213 RLEKDIINLIMLRQIWCESENLR----LLYQFLV----------VEYQNKSA--NSVLLN 1256
Db   1259 KIFNFLFEIITGKNIYYQINDIHNIVALFKEFKIYDCKHVLKKSIKVLNKYAKKNSKEFS 1318

Qy   1257 VSSNNGDIIL-LN---------KKLVQDNIKNS----------MDHNNIH 1286
Db   1319 LFENNQHVVLDINKHMLFNDDEKKLTTCNIKQNEQEQIKTKVLYDHDNIN 1368

RESULT 29
AAB18311
ID   AAB18311 standard; Protein; 807 AA.
XX
AC   AAB18311;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:169.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
```

```
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 389-391; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   807 AA;

  Query Match             10.2%;  Score 696;  DB 21;  Length 807;
  Best Local Similarity   26.6%;  Pred. No. 3.6e-31;
  Matches  253;  Conservative  151;  Mismatches  300;  Indels  246;  Gaps   43;

Qy        328 SSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNHLLSNNNNNKIN 387
              ||: || :|      ||: ||:|             | | |  ::| | |  :  :||:| || |
Db         10 SSKEKNEEN--DIINKCDDSN-------KINGKE-NIFAVEKVGINESGHMSNDNINK-N 58

Qy        388 RGKINDKIERNNILKNKSFDKPREGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYN- 446
              : |    | :: |  |    : :           :|  | :   ||:|  :  ::  :|| |:   |
Db         59 QEKNKKKKKKKNTHKKVNINNTHINIHTTNDKNNG-QDINKPEVIERDNIINIKNDTNNI 117

Qy        447 -DNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNN-----MINHMINNNI----- 495
               |:: |:: | || |: :||:|||: | |||| ::|  ||          |   :  |:|
Db        118 LDSSYNEEGNENNRNDINNNNNNNNINNNNNINNSCSNNYGLKKKITLLKRNDIKDEGY 177

Qy        496 -NNNINVNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKK 554
               | ||   || ||  |||| ::|                         |||:
Db        178 NNENITTLNNKNNLKNNNNYNDNR-----------------------NNNN-------- 205

Qy        555 NLRDNIKLKNRSSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKME 614
              | ::||  | ::|                     |: |:  : : ||      || || :
Db        206 NNKNNINNNNNNNC------------------CSEKTLEQREKEYNKIRARIFSNFNKKQ 247

Qy        615 DNTPKDILYESRNVSN--MNGNVLLGLNKNTHHDISTKDENHNDNK---INNGVINIINN 669
              |  |    |  |::: :| |:: :             |: ||:   |||    :| :|
Db        248 KNVQKT---EQNNLNHTYLNNNIINNI-------------NNGDNQYAYINN-FYHIYHN 290

Qy        670 SNVNSINNSNMNSINNSNMNSNSIYKSN----YNSNQSISDVQIRYVNEM-DTSNKNNDN 724
              :: | |   |   | | | :: :| | |      |: |        | | | |  | ||
Db        291 NSYNHIYRQNNIPICNINNHAPNIEKLNNPYYYHDN------HIAYTNYMYSTQNKM--- 341

Qy        725 IFFDAISCDNNM---------YPNITNNNNNNNNNNNNNNNIDVENYNNRDGTNNSMKL 775
                         |||           |  ||||||| |||||||:  | ||     |  : :
Db        342 ---------NNMKTKQIGHYGINNEDNNNNNNNINNNNNNNINNNNINN---NNVPLCI 389

Qy        776 YAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYNLNDKNYLID-LNNKE- 833
                 :::|  :  : |: |:|        | |       | |  :  |   :|:|: |  |: |||
Db        390 PQLDNYNKTKNNFNQGTNNFNQGTNNFNKCTNNFNNAKNHIKHNINNTNKNIEHLNNHSI 449

Qy        834 -----QKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCG 888
                        ::|||:  : |:| |    : :  |   |   |  |    :       | :||: ::
Db        450 YNFVYPENKNIYDANGNLINNNISYTQLKMNNNINFNIHMESPINQQHNNTFKV----N 504

Qy        889 SHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNN--IYNNINDDDTFQND 946
              : |:  |    |:||   |:  |  |:  |  :||||         |||  :| :||     |||
Db        505 NDTNFFNEPTNKMKK---KNKEKKNIHFNNNNNN-------NNNKCLYKDIN-----QND 549

Qy        947 YCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNED 1006
              :                 ||: ||:|  || | |  :            ||  :  ::|:   :  : :
Db        550 H-----------NNSIINTN--QNFDHINNV-------KNTEQNLQKKHNKMSQVSKQS 588

Qy       1007 IYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNV---DNNNVDNNVDN------- 1056
                      |::         :  :  |  |||:||  | ::  ||    ||      :: ||
Db        589 NNKNNKNNSHLK----KQINTNTNNNMDNKNNSHISKNVIVDDNKLKSSHADNSNEIVTK 644

Qy       1057 ---------------NDKNNVDN----NNVDNDDDDVDFHNIKNFNNNEY-----LSYF 1091
                                 |  |||:|    ||::: :: :   :|: | ||   |      :
Db        645 GKKKKNTNKKKKINNINSVNNVNNINSMNNINSMNNIISMNNVNNMNNPMYFPNVNIQKD 704

Qy       1092 QKNVDTIINNCLNSLDISSM---YDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSL 1148
                    |:  : ||   | :| ::     :  :   | |||: :          :|:      | :: | |
Db        705 DSNIALLYNNKPN-IDFNNFQLNHINNHMIQNNIMTNNVML-NNNLTTSNFNYNLINYSY 762

Qy       1149 E---------EIDKTAQ-SIYEKRKVLLTKLLLLFKKNVDTQINNETSDL 1188
               |            ::|     |:||||            :  :  :  | |::  |:
Db        763 EPFYEENLMNDLDYCRDISLYEKR----------YDRGDNLQQNHKRYDI 802

RESULT 30
ABB59733
ID   ABB59733 standard; Protein; 626 AA.
XX
AC   ABB59733;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 5991.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03836.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
```

```
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 5991; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   626 AA;

  Query Match            10.0%;  Score 686.5;  DB 22;  Length 626;
  Best Local Similarity  42.1%;  Pred. No. 9.2e-31;
  Matches  159;  Conservative   67;  Mismatches  131;  Indels   21;  Gaps    7;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
                :| | |||||    |     :: |:::: :   || : |||   |:: |::| | |   | |||:
Db        187 QIMVCVRKRPLRRKELADREQDVVSIPSKHTLVVHEPRKHVNLVKFLENHSFRFDYVFDE 246

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGSQP-YGQSDTPGIFQY 122
                  | |||| | :|||   ::: | : :||||||||||||||| |   |    ||    ||:
Db        247 ECSNATVYEFTARPLIKHIFDGG-MATCFAYGQTGSGKTYTMGGQFPGRHQSSMDGIYAM 305

Qy        123 AAGDIFTFLNI--YDKDNTKGIFISFYEIYCGKLYDLLQKRK-MVAALENGKKEVVVKDL 179
              || |:|: |       |:| | | :: ||:|| :::|||    |  :   ||:  ::| |  |
Db        306 AAKDVFSTLKTVPYNKLNLK-VYCSFFEIYGTRVFDLLMPGKPQLRVLEDRNQQVQVVGL 364

Qy        180 KILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                    |     |:: :   |  :|   |   | | :|||||:   | |:         || : ||
Db        365 TQNPVQNTAEVLDLLELGNSVRTSGHTSANSKSSRSHAVFQIVLRSAAGEKLHGKFSLID 424

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:|||||  | ::||: :|: ||:||| ||||||:      :|:||| |:||:|||| |
Db        425 LAGNERGADNSSADRQTRLEGSEINKSLLVLKECIRALGRQSSHLPFRGSKLTQVLRDSF 484

Qy        300 VG--KSKSIMIANISPTISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTERISILDS 357
              :|   | |: ||| |||  :     | |||||||: |||         :|        ||
Db        485 IGGKKVKTCMIAMISPCLHSVEHTLNTLRYADRVKELSVESIPSKRMPDAN--------- 535

Qy        358 KGSEMNASSIENVVIKSN 375
                   : ::|: ::| :|:
Db        536 ----LGSTSMSDIVCQSS 549


RESULT 31
AAO22566
ID   AAO22566 standard; Protein; 2150 AA.
XX
AC   AAO22566;
XX
DT   28-OCT-2002  (first entry)
XX
DE   Wooden leg (WOL) gene related protein SEQ ID No 17.
XX
KW   Wooden leg; WOL; vasculature; transgenic plant; agronomic; longer root;
KW   wood production; plant; promoter; tree; crop plant.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200244337-A2.
XX
PD   06-JUN-2002.
XX
PF   29-NOV-2001; 2001WO-US45053.
XX
PR   29-NOV-2000; 2000US-253739P.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
PA   (HELA/) HELARIUTTA Y.
PA   (MAHO/) MAHONEN A P.
PA   (BONK/) BONKE A W M.
PA   (KAUP/) KAUPPINEN L.
PA   (RIIK/) RIIKONEN M.
XX
PI   Helariutta Y,  Mahonen AP,  Bonke'AWM,  Kauppinen L,  Riikonen M;
PI   Benfey PN;
XX
DR   WPI; 2002-599423/64.
XX
PT   Novel isolated polypeptide (WOODEN LEG) with ability to regulate a set
PT   of asymmetric cell divisions that establish vascular tissue in root and
PT   hypocotyl development, useful for improving agronomically valuable
PT   plants -
XX
PS   Example 1; Page 151-157; 187pp; English.
XX
CC   The invention relates to an isolated WOODEN LEG (WOL) polypeptide,
CC   comprising 15 contiguous amino acids of a fully defined Arabidopsis
CC   WOODEN LEG protein sequence of 1057 amino acids as given in the
CC   specification, and to its encoding nucleic acid. The invention also
CC   relates to an amino acid sequence of domains of protein, e.g., N-terminal
CC   region, C-terminal domain, etc; or is a naturally occurring allelic
CC   variant of the above mentioned polypeptide sequence. Expression levels of
CC   the nucleic acid can be modified to improve the vasculature in transgenic
CC   plants and enhance the agronomic properties of such plants. Also the WOL
CC   promoter is used to drive expression of a heterologous coding sequence of
CC   trees to improve wood production. The WOL nucleic acid may be used as a
CC   molecular marker for a qualitative trait loci, e.g., longer roots or
CC   enhanced wood production, in molecular breeding of crop plants. The
CC   nucleic acid is also useful in DNA amplification assays to identify the
CC   endogenous WOL genes, WOL mutant alleles and/or WOL expression products
CC   in cultivars as compared to wild-type plants. They can also be used as
CC   markers for linkage analysis of qualitative trait loci. The WOL protein
CC   and/or antibodies can be used as diagnostic reagents in immunoassays to
CC   detect expression of the WOL gene in cultivars and wild-type plants. The
CC   WOL protein, its encoding nucleic acid, and its corresponding antibody
CC   are useful for improving agronomically valuable plants e.g., trees. This
CC   sequence represents a protein relating to the wooden leg (WOL) protein of
CC   the invention.
XX
SQ   Sequence   2150 AA;

  Query Match             9.9%;  Score 676;  DB 23;  Length 2150;
  Best Local Similarity   24.5%;  Pred. No. 1.5e-29;
  Matches  242;  Conservative  146;  Mismatches  255;  Indels  344;  Gaps   41;

Qy        416 TFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINKKHNNNNNNNNDNNDNNNDNNNNN 475
              ::| :| ::::           :|        |:|  |   :|||||||:|| |||:|||||
Db         85 SYGNHSFIHNVSPTSP-------SYDINNNNNNNNNNNNNNNNNNNNNNSNNNNNNNNN 137

Qy        476 N--------------------------------------------NDSSSM--VNNMINH 489
              |                                            |:|||:  :|  :|
Db        138 NKNNYNNNYYYSPIENSNISKSLEESVLNQFPHNFNLNSSNNNYLNNSSSLHNINQSVNS 197

Qy        490 MINNNINNN----INVNNNNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNN 545
              : ||| |        || ||||||||||||:|::                     | : |||
Db        198 LSNNNNNQTNQQPINNNNNNNNNNNNNSNNSN--------------------NSNNNNN 237

Qy        546 DKSIFLHKKNLRDN-IKLKNRSSCDNIMNKKKNNLHLARHSVGSKL---TMFSYDPQKNK 601
                   : |: |:  | |  |: :  :        | |  |: | :      :|:   |||
Db        238 GN----NNNNITDSPTKSKRHSTYETNIGS-----HQRRKSIQSLIANSAIHSFSKLKNK 288

Qy        602 DNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINN 661
                        : :||  :                |     :| |:::        |:|:|  :
Db        289 ----------PLSSSTPSTV-----------NTCGAVNNNSNN-------NNNNNNST 319

Qy        662 GVINIIN-----NSNVNSI-------NNS-------------------NMNSINNS-NM 688
              | :  |        : |:|:|          |||                   |::|:|:  |:
Db        320 GSLGAIPMDRSFDGNINTITEESTGGNNSPRSNCGSNCGSNGGIPLSPRNLSSLNSGVNV 379
```

```
Qy        689 NSNSIYKSNYNSNQS-ISDVQIRYVN-EMDTSNKNNDNIFFDAISCDNNMYPNITNNNNN 746
              :  :|: :| |:| | :   :  |::|  :: || ||:|          :||: ||   ||:|
Db        380 SPRNIHLNNLNNNSSNLPPLSPRHINFHINVSNLNNNN--------NNNINPNNNPNNSN 431

Qy        747 NNNNN---NNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKN 803
              |:|||    |||:||       |       :||                |  |  | :|| |
Db        432 NSNNNVSPRNNNHNISPRGSNISPRSNN-------GGSTTISPRNISNNNNIIN------ 478

Qy        804 NQDGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCD---------NNIIQNRNDFE 854
                               |:|: | |    |  : : |::  :          |: :   :
Db        479 ---------------NINNNNILTPPRNSPRLE-NVNPTNSPRLLATSLNSTLPIVSSLT 522

Qy        855 KKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDN 914
                     |  |||      :|||  | |        | |      |   : :  |   :  |   :|
Db        523 SSNNNNQSNNNTNPSINNNNGRN-------GHCIQTISEEILGNKPVVYNNGNNN----- 570

Qy        915 HIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQ---- 970
                  ||||:::|::||||  | |::                          || |||:|:
Db        571 ----NNNNTNNSTTSNNNITTNNNNN-------------------NNNNINNNVLSTPRK 607

Qy        971 ----NDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYD 1026
                      |       ||| |: ::               :::  :|   :     |   :|   ::|  |
Db        608 RTKGNHS---KTNSLQDFETS-----------SMNGGDDSIGAGSGGSLRRRNKDDND 652

Qy       1027 DND-----------NNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDNNNVDNDDDD 1073
              :||              ||| :||: || :|| :||| :|| :||: || :||| :|:::
Db        653 ENDGNSNNTNSNNSNNNNNNNNNNSSNNNNNNSNNNNNNNNNNNNNNNNNNNNNNNNN 712

Qy       1074 VDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEKDN 1133
               : :|   |  |||                 ||  |    :      :        |:  :    : | |
Db        713 NNNNNNNNNNNNN------------NNYHNGATMMMSH-------NHQSIGMSSSPKKN 752

Qy       1134 VIKKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLV 1193
                |  :       :| ||  : :  | :|                :| ||           |    :   :  ||
Db        753 NFKPF----SRNCSLMGMGRRAWAI----------ILGLF-------IVGSSISILATLV 791

Qy       1194 MCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNK----S 1249
              : :     |:  | |       :|: :|      |||                    :|: |:
Db        792 LRYSEENSIADDF-----ARVARD--RFTMLR-----------------IEFNNRLYITQ 827

Qy       1250 ANSVLLNV-SSNNGDIILLNKKLVQDN 1275
                 |:||:|  | : |   :    ||  ||
Db        828 TLSLLLSVFPSTSEDQFVPFSKLWSDN 854
```

RESULT 32
AAB56496
```
ID   AAB56496 standard; Protein; 410 AA.
XX
AC   AAB56496;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human prostate cancer antigen protein sequence SEQ ID NO:1074.
XX
KW   Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW   neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW   vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW   antibacterial; gene therapy; neural; immune; reproductive; renal;
KW   gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW   wound; infectious disease.
XX
OS   Homo sapiens.
XX
PN   WO200055174-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05988.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587513/55.
DR   N-PSDB; AAF15699.
XX
PT   Prostate cancer associated gene sequences, referred to as prostate
PT   cancer antigens, useful for treatment, prevention, and diagnosis of
PT   disorders such as prostate cancer -
XX
PS   Claim 11; Page 1498-1499; 2338pp; English.
XX
CC   AAF15566 to AAF16505 encode the human prostate cancer associated
CC   proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC   The prostate cancer antigens can have neuroprotective, cytostatic,
CC   cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC   nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC   and can be used in gene therapy. The prostate cancer antigen
CC   polynucleotides may be used for detection of prostate cancer, chromosome
CC   identification, as chromosome markers, and for numerous other diagnostic
CC   or research purposes. The prostate cancer antigens may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC   AAB57303 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence   410 AA;

  Query Match             9.8%;  Score 672;  DB 21;  Length 410;
  Best Local Similarity   46.6%;  Pred. No. 3.7e-30;
  Matches  149;  Conservative   59;  Mismatches  102;  Indels   10;  Gaps    8;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
              :| | ||||||:: |  ||: |:|:: : | | : ||: |||:|||:|   |   |  ||:
Db         57 RICVCVRKRPLNKQELAKKEIDVISIPSKCLLLVHEPKLKVDLTKYLENQAFCFDFAFDE 116

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
               |  |  ||  | :||:  ::| |   :||||||||||||:|| |         |: :  ||:
Db        117 TASNEVVYRFTARPLVQTIFEGG-KATCFAYGQTGSGKTHTMGGDLSGKAQNASKGIYAM 175

Qy        123 AAGDIFTFLN--IYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
               |  |:|   |   | |    :  ::::|:||  |:||:|| |:  :    ||:|:||:|  |   |:
Db        176 AXRDVFLLKNQPCYRKLGLE-VYVTFFEIYNGKLFDLLNKKAKLRVLEDGKQQVQVVGLQ 234

Qy        181 ILRVLTKEELILKMID-GVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFID 239
                 | : :: ::||  | |   |  |    |  ||||||   | |:    |     || : :|
Db        235 EHLVNSADD-VIKMXDMGSACRTSGQTFANSNSSRSHACFQIILR--AKGRMHGKFSLVD 291

Qy        240 LAGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIF 299
              |||:|||| | | ::||: :|| ||:|||||||||||:   :|  | |||:|:||:|||| |
Db        292 LAGNERGAXTSSADRQTRMEGAEINKSLLALKECIRALGQNKAHTPFRESKLTQVLRDSF 351

Qy        300 VGK-SKSIMIANISPTISCC 318
              :|: |:: ||| ||| || |
Db        352 IGENSRTCMIATISPGISSC 371
```

RESULT 33
ABB05593
```
ID   ABB05593 standard; Protein; 604 AA.
XX
AC   ABB05593;
XX
DT   22-APR-2002  (first entry)
XX
DE   Mitotic centromere-associated kinesin NYD-KIF2 protein.
XX
KW   Human; mitotic centromere-associated kinesin; NYD-KIF2;
KW   human genome technology.
XX
```

```
OS   Homo sapiens.
XX
PN   CN1318561-A.
XX
PD   24-OCT-2001.
XX
PF   11-APR-2001; 2001CN-0113509.
XX
PR   11-APR-2001; 2001CN-0113509.
XX
PA   (UYNA-) UNIV NANJING MEDICAL.
XX
PI   Sha J,  Zhou Z,  Li J;
XX
DR   WPI; 2002-122927/17.
DR   N-PSDB; ABA93407.
XX
PT   Human testicular mitosis dynein gene encoded protein for immunizing an
PT   animal and preparing monoclonal and polyclonal antibodies  -
XX
PS   Disclosure; Page 2 (Disclosure); 6pp; Chinese.
XX
CC   The present invention belongs to the field of human genome technology.
CC   The present invention describes the human testis mitotic centromere-
CC   associated kinesin NYD-YIF2 gene which has a cDNA sequence of 2334 base
CC   pairs (bp) containing an open reading frame sequence of 1815 bp from
CC   position 135 to 2156, encoding a sequence of 604 amino acids and has a
CC   Genbank number of AF333335. The present invention also describes:
CC   (1) utilising the NYD-KIF2 gene to prepare a fusion protein;
CC   (2) utilising the protein to immunise an animal and to prepare
CC   monoclonal and polyclonal antibodies; and (3) utilising the NYD-KIF2
CC   gene in preparing a testicular specific function gene expressing chip.
CC   The expressed protein may become gene medicine for treating related
CC   disease. The present sequence represents the human testis mitotic
CC   centromere-associated kinesin NYD-YIF2 protein from the present
CC   invention.
XX
SQ   Sequence   604 AA;

  Query Match             9.7%;  Score 663;  DB 23;  Length 604;
  Best Local Similarity   44.7%;  Pred. No. 1.9e-29;
  Matches  149;  Conservative   61;  Mismatches  115;  Indels    8;  Gaps    6;

Qy          4 KIKVVVRKRPLSELEKKKKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDD 63
Db        213 RICVCVRKRPLNQRETTLKDLDIITVPSDNVVMVHESKQKVDLTRYLQNQTFCFDHAFDD 272

Qy         64 TVDNFTVYENTIKPLIIDLYENGCVCSCFAYGQTGSGKTYTMLGS-QPYGQSDTPGIFQY 122
Db        273 KASNELVYQFTAQPLVESIFRKG-MATCFAYGQTRSGKTYTVGGDFSGTAQDCSKGIYAL 331

Qy        123 AAGDIFTFL--NIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLK 180
Db        332 VAQDVFLLLRNSTYEKLDLK-VYGTFFEIYGGKVYDLLNWKKKLQVLEDGNQQIQVVGLQ 390

Qy        181 ILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNIDLKDINKNTSLGKIAFIDL 240
Db        391 EKEVCCVEEVLNLVEIGNSCRTSRQTSVNAHSSRSHAVFQIILK--SGGIMHGKFSLVDL 448

Qy        241 AGSERGADTVSQNKQTQTDGANINRSLLALKECIRAMDSDKNHIPFRDSELTKVLRDIFV 300
Db        449 AGNERGADTTKASRKRQLEGAEINKSLLALKECILALGQNKPHTPFRASKLALVLRDSFI 508

Qy        301 GK-SKSIMIANISPTISCCEQTLNTLRYSSRVK 332
Db        509 GQNSSTCMIATISPGMTSCENTLNTLRYANRVK 541

RESULT 34
AAP60452
ID   AAP60452 standard; Protein; 537 AA.
XX
AC   AAP60452;
XX
DT   25-MAR-2003  (updated)
DT   27-JUN-1991  (first entry)
XX
DE   Sequence of the Asparagine-Rich Protein (ARP)
DE   Ag319 of Plasmodium falciparum.
XX
KW   Malaria vaccine.
XX
OS   Plasmodium falciparum.
XX
PN   WO8606075-A.
XX
PD   23-OCT-1986.
XX
PF   11-APR-1986;   86WO-AU00092.
XX
PR   25-JUL-1985;   85AU-0001640.
PR   11-APR-1985;   85AU-0000108.
PR   11-APR-1985;   85AU-0001008.
PR   11-APR-1986;   86AU-0056037.
XX
PA   (HALL-) HALL INST MEDICAL RES W & E.
PA   (KEMP/) KEMP D J.
XX
PI   Kemp D,  Anders R,  Brown G,  Coppel RL;
XX
DR   WPI; 1986-291646/44.
DR   N-PSDB; AAN60392.
XX
PT   New DNA molecules encoding Plasmodium falciparum antigens -
PT   namely the SHARP ARP and MESA antigens
XX
PS   Claim 16; Fig 8; 52pp; English.
XX
CC   The antigens of the invention are prod. using recombinant DNA
CC   techniques. They may be used in vaccine compsns. to stimulate an
CC   immune response against P.falciparum, the cause of human malaria.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence   537 AA;

  Query Match             8.8%;  Score 604.5;  DB 7;  Length 537;
  Best Local Similarity   29.1%;  Pred. No. 3.2e-26;
  Matches  209;  Conservative   80;  Mismatches  193;  Indels  235;  Gaps   40;

Qy        454 HNNNNNNNNDNNDNN----NDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNN 509
Db          1 YNNNNKNNNNNDDGNMNFKNQYNNNYKFDEEEHLR----------NNSIDMNNSNENEN 49

Qy        510 NNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCD 569
Db         50 ENENK--------NY-HTGGSSMKNTD------NNKTDTSYLR------NI--------- 79

Qy        570 NIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVS 629
Db         80 NNINEYK---------GSAKT------QNNND------NMNTNEDNN-----YQNNNMN 112

Qy        630 NMNG-NVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINNSN- 687
Db        113 RNNSINIKMG----------SQDKNQNSN--NNFYINY---QNTNEFKDNKKNNMNNSNT 157

Qy        688 ---MNSNSIYKSNYNSNQS-----ISDVQIRYVNEMDTSNK-----NNDNIFFDAISCDN 734
Db        158 MHSRNSNVI--NNYTNQQTRFSSFMMNNNIHFKNKYDNNNNMKGTINNDNNNMDY----N 211

Qy        735 NMYPNITNNNN---NNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDN--- 788
Db        212 KFYTNYMNKNNLKFNNNNNNNNNNGVFSNRNLNNNNNINNNMNKMNYNYQNRKNSMNNNM 271

Qy        789 NKNTSNIQNINTNKNNQDGNVNYSMN-----FCHYNLNDKNYLIDLNNKEQKDKNIHGCD 843
Db        272 NNNMNNNMNHNMNNNNMNHNMNNNMNNINMSMMNYNNNESN----TANRENNNMNNYGYD 327
```

```
Qy        844 --------NNIIQNRNDF-EKKKKTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSID 894
                      || : | |:  |  :  |  :||:  | ||| |||:  :| | |    ||
Db        328 DNTAVGYNNNYLNNNNNMNENSENKNVADNNDSSINNNNTLNNNN---EYNL-GEDKYID 383

Qy        895 -------NMKNNEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNIN--DDDTFQN 945
                     : | |: ||  | |:| ||             : |||: :|:| | |  |
Db        384 INSIMKKSKKGNDGKNK--KKNKMNMNH-----------NMNNNMNHNMNSLDSDMSPN 429

Qy        946 DYCH-NDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKEKYTYPTLSTN 1004
               : |  |      : || |:  :: |                           | ||
Db        430 YHAHVKPNQMNFEQTNNDNMKVHVNNN----------------------------TPST- 460

Qy       1005 EDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNNDKNNVDN 1064
               | :::                  |: |:||: |   ||:  || || :| :  |   |
Db        461 -DFFSR------------------NSAVNNNSSNG--NNMKLNNNKNNNNNINMNENQN 498

Qy       1065 NNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSL--DISSMYDDTKEIL 1119
              || | ||||                 : :|||         || |  |:: :  |  :|
Db        499 NNEDEDDDD----------------WGEKNV------ILNQLEADLNDLKKDDLFVL 533
```

RESULT 35
AAE29358
```
ID   AAE29358 standard; Protein; 1331 AA.
XX
AC   AAE29358;
XX
DT   27-JAN-2003  (first entry)
XX
DE   Plasmodium falciparum BBP-5 protein.
XX
KW   Band 3 polypeptide; malarial infection; drug resistance; vaccine;
KW   protozoacide; gene therapy; BBP-5.
XX
OS   Plasmodium falciparum.
XX
PN   WO200270542-A2.
XX
PD   12-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-US06415.
XX
PR   02-MAR-2001; 2001US-272930P.
XX
PA   (SELI-) ST ELIZABETH'S MEDICAL CENT INC.
XX
PI   Chishti AH,  Oh SS,  Liu D,  Goel V;
XX
DR   WPI; 2002-759814/82.
DR   N-PSDB; AAD47004.
XX
PT   New isolated Band 3 polypeptide which selectively binds to merozite
PT   surface protein-1, useful for the prevention and treatment of malarial
PT   infection  -
XX
PS   Claim 27; Page 148-153; 163pp; English.
XX
CC   The invention relates to an isolated Band 3 polypeptide that comprises
CC   any of 4 20 residue amino acid sequences, or their fragments that bind
CC   to an MSP-1 polypeptide or a polypeptide with any of 8 291-1331 base
CC   pair sequences. The methods and compositions of the present invention
CC   are useful for the prevention and treatment of malarial infection. The
CC   present invention, develops new and more improved methods based upon
CC   inhibiting the particular interactions between the malarial parasite and
CC   a cognate molecule present in the host and subsequently minimising
CC   harmful side effects and drug resistance that may be due to non-specific
CC   therapeutic approaches. The invention is useful in gene therapy. The
CC   present invention also provides a vaccine for malaria. The present
CC   sequence is Plasmodium falciparum BBP-5 protein.
XX
SQ   Sequence   1331 AA;

  Query Match            8.7%;  Score 596;  DB 23;  Length 1331;
  Best Local Similarity  22.3%;  Pred. No. 2.7e-25;
  Matches  317;  Conservative  180;  Mismatches  412;  Indels  512;  Gaps   68;

Qy        122 YAAGDIFTFLNIYDKDNTK------GIFISFY---EIYCGKLYDLLQKRKMVAA------ 166
               | : |:  ||  |         :  ||    |:| |  ||:   :| |
Db         52 YVSSDLLNFLKCYSNLNINLNKVPYDLVYSFLLDGELYLG--YDISVFILLVKAEHFEYC 109

Qy        167 ----LENGKKEVVVKDLKILRVLTKEELILKMIDGVLLRKIGVNSQNDESSRSHAILNID 222
                  || |:   :        || :  :|           :|| |:      :|  :
Db        110 RRIDNENSDKKESFR--------TKNKSTIKR-----------SSQIDDEDNLQGLLIKE 150

Qy        223 LKD----INKNTSLGKIAFIDLAGSERGADT---VSQNKQTQTDGANINRSLLALKECIR 275
               :|     :|:|      |        ||:  :    : ::|     |   :    |    |
Db        151 KEDYLSFLNENNE----ALKQYMESEKRGNPLWHLDESKYMDKDWYDEEDSSFIFKPTFN 206

Qy        276 AM-------DSDKNHIPFRDSELTKVLRD-----IFVGKSKSIMIANISPTISCCEQTLN 323
               :        ::  |:  |  :  :  :  |          ||  |  : :  |  :    | :|:
Db        207 YLGKNNNNNNNHNNNNAFSNFVMGNLSSDNISGCFFVEKLNAYLFAMLD---KCSNKTVI 263

Qy        324 TL-------RYSSRVKNFKNKSTCINEEDDTNTERISILDSKGSEMNASSIENVVIKSNH 376
              ::          |: ||      |    :: :|          :  |  |: |:   ||
Db        264 SVFPYEKFGRHESR-----NLAIQFSQYEDYMHRIIEDRLYANIQNNLPSVHNMKNMSN- 317

Qy        377 LLSN-NNNNK---INRGKINDKIERN-----NILKNKSFDKPREGFTSTFGKYSSLNDID 427
              ::| ||||| |||  |::    ::         |||     :|| :          | :|  :|
Db        318 -MNNINNNNKDIIINRSGISNGNSQSVPCFENIL---DYDKLK------FVEY--INSFS 365

Qy        428 KIKKNKKKGLINYKSTLYNDNTINKK----HNNNNNNNNDNN-----------DNNNDNN 472
               :||:    :|    :|         |        ||||  :   :|               :|: |
Db        366 DVKKSSSFDIIGSSKNIYEQGENLKNYCIYHNNNFESGFENYILENKQPLELIENHFDIM 425

Qy        473 NN------NNNDSSSMVNNMINHMINNNIN-NNINVNNNNNNNNNNNNSHNNHLPQPNYA 525
               |      | |      ||:  :  :| | | | | :  :|:|||  : |  | :      |
Db        426 ENIKGMYDNTNQEEMNFNNVSGLLREDNSNMNEIYLTRDNHNNNYHENEENIYSINIKY- 484

Qy        526 FTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDNIMNKKKNNLHLARHS 585
                            :| | || |    |  ||:: :| : |  ||     |   || |  :
Db        485 ------------INNHFNNKDDMI-MKCKNMKGSISMDNNSS-----NSNSNNTHFEK-- 524

Qy        586 VGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHH 645
                   |: | :|     |:  | | ::| |: |   |    |      :      |:|  : |:|:
Db        525 -----TLESINPD---DHNIFNSEMDSMK-NENND---EEEQTATSIYNILGKIGKDTY- 571

Qy        646 DISTKDENHN-DNKINNGVINIINNSNVNSINNSNMNSINNSNMNSNSIYKSNYNSNQSI 704
               |      |:| || |||     | |: |  |  |  | :: ||:  |||    |:|||
Db        572 -IKRCSSNYNYDN--NNG---YSNESSDNYNNGYNDSTDNNNGYNSN----SSYNSNN-- 619

Qy        705 SDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNIDVENYN 764
                          || | :| ||::         :|||   |  | |||||  ||||||      |
Db        620 --------NEDDNNNNNNND-----ENCDNNNNHNNNNYNNNNNYGNNNNNNN------N 660

Qy        765 NRDGTNNSMKLYAYNSHN-------LFQPDNNKNTSNIQNINTNKNNQDGN-----VNYS 812
              |:| ||      : |::|             |:| |  | || |:: |: |:  :|:       | |
Db        661 NKDNNNNDGNGSSNNNNNDDDDEEEEDDEDDNNNNNDDDNMSDNEEMEDNDEDNDEYNNS 720

Qy        813 MNFCHYNLNDKNYLIDLN---------NKEQKDKNI-HGCDNNIIQNRNDFEKKKKTNFY 862
                :    |  | |:  ||            | :|||  |  |   |   |  : ||   :
Db        721 NDSYKYEEKDSNHEKDLKKDIIEGDMINSVKYDKNIGHHTTNKSEISTNYFENSCNMSVN 780

Qy        863 NNNN-----------------IVIVNNNMGNNNSPRMKYGLC-------GSHTSIDNMKN 898
              |:||                  : : | |: ||           |           ||  :|| |
Db        781 NSNNEAYDDNCNNGFMNHDEGLTLNNGNVSNNK--------CDIIIPEDGS-VMYENMIN 831

Qy        899 NEMKNNEMKDNEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDY---------- 947
                 : | :  |   :|:: :||:| | ::::|:||||:  : | |
Db        832 ---RGNGLTSNINNNNNV---SNNNSISCNADDNVYNNINN---YINTYMETTNNKNHIE 882

Qy        948 --CHNDNTFTIRRK-NNTNINSNIYQNDDIIYTINSLNDYMSNTLLHFKE---------- 994
                 |: |:  |     :| :||     | |:|  | |:  |:  |  | ||
Db        883 NRCNQDSYSTNEEPLSNHSINDPGKIKDGIMYDGNDLD--MNGTQEHSKEEGMDVFEPNF 940

Qy        995 ------------KYTYPTLSTNEDIYNKEMEGKHIR----LDDQDK-------------- 1024
```

```
                  |:  |      :  |  |  : ||  |      |||: |
Db        941 FELKRNSSDGQNKHLEP--GVQKKISKKRSKVKHERNSKILDDEKKEVLNKVSQITRVGG 998

Qy       1025 ------------------------------------------------------------ 1024

Db        999 VCFDKNRQRWIAHWKIDGKYHKHYFPISQYGFENARERAVSCRKQAEKLFNLPEIQPRNR 1058

Qy       1025 -------------------------YDD-------------------------------- 1027
                                       ||:
Db       1059 WNQIKVNGTSHIKKAAKLPRCEGIGYDELSQSWVSTFVVHKKFSIEELGFYEAREKAIYC 1118

Qy       1028 --------------------------------------------------NDNNNVDNN 1036
                                                                ::|| | :|
Db       1119 RKTFEKVNVHDDYECLLNDRLGLRNEEKDELSDLINIDKNALDNLELETSVHNNNKVKHN 1178

Qy       1037 NKNNVDNNVDNNNVDNNVDNNDKNNVDNN--NVDNDDDDVDFHNIKNFNNNEYLSYFQKN 1094
               | ||  || :||| :|| :|::|  : ||   :|||::::|     ||  :|::||   |:
Db       1179 NNNN--NNNNNNNNNNNNNNSEKMRIKNNDFSVDNNNENVGTGEIK-ISNDKYLKITQEA 1235

Qy       1095 VDTIINNCLN-SL-DISSMYDDTKEILN-NILLSKYKAEKDNVIKKYINEDIKNMSLEEI 1151
              :: |::|  : || :|      | ::  | | || |:           |:|   :||:|
Db       1236 IEMILSNIKHKSLPEIKMKLIDKQKFENYNTLLDKH--------FKFIT-SVKNIS---- 1282

Qy       1152 DKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDL 1192
                          | : : || | :         ||| |
Db       1283 ------------QLRRYISLFHKFIIYHTLPHNISLRKQL 1310
```

RESULT 36
AAB18172
ID   AAB18172 standard; Protein; 2485 AA.
XX
AC   AAB18172;
XX
DT   07-NOV-2000 (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:29.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 75-82; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   2485 AA;

```
  Query Match             8.6%;  Score 587;  DB 21;  Length 2485;
  Best Local Similarity   20.5%;  Pred. No. 1.8e-24;
  Matches 356;  Conservative  249;  Mismatches  517;  Indels  616;  Gaps   85;

Qy         21 KKDSDIITVKNNCTLYIDEPRYKVDMTKYIERHEFIVDKVFDDTVDNFTVYENTIKPLII 80
              |:: :||: | :|  : ||     :  |||  :  :|::            :  |:: | |
Db        359 KENINIISNKESC--HKDEK----EEEKYIMYNSNLVEE-----KKQKKMIWNSLNVLPI 407

Qy         81 D-LYENG--------CVCSCFAYGQTGSGKTYTMLGSQPYGQSDTPGIFQYAAGDIFTFL 131
              | | :||        |       ::       |:  :  :: | |:
Db        408 DILLKNGHDEINKEICKKKKKSFFSQNDIKSKMLYNNKSYSKSEKV------------- 453

Qy        132 NIYDKDNTKGIFISFYEIYCGKLYDLLQKRKMVAALENGKKEVVVKDLKILRVLTKEEL- 190
               :|  :      ||   :   :  |:  :   :  :  : | || | : | ||   :  :|
Db        454 -LYTNNKNSNTFIPIF--FLNKVGDKFKNSENIYDMYNNKKNVYIHDKKIYTNMYSNKLK 510

Qy        191 ----------------ILKMIDGVL----------------------------LRKIG 204
                              | |::|  |                             ||
Db        511 QKHYYSTSNINLLYNNIGKVLDNGLHLSNNMYCRLNSNPPYKSISLINNNVFFYKKRKSN 570

Qy        205 VNSQND----------ESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNK 254
               |: |:           | ::| |:|                  ||:  ::   || |:  :|
Db        571 SNNNNNNNNISSSSSSSSKKNHVIIN-----------KKISSYNIHYKER-KDSFKEN- 616

Qy        255 QTQTDGANINRSLLALKECIRAMDSDKNHIPFFRDSELTKVLRDIFVGKSKSIMIA----N 310
                         | ||  : : | |:   | : :      :|:|   :: |          |
Db        617 ------------FLFFKE--KILPSKKDTCVFNERQ-----KDLFEKSNEHIKCVSSFNN 657

Qy        311 ISPTISCCEQT----------LNTLRYSSRVKNF---KNKSTCINEEDDTNTERISILDS 357
               |  ||                |::|:  :  |      ||    |    |    |:  :|
Db        658 TSDDISSHSSVNKKEPFFALKNNSIRHIPKENNIIYTSGKS--FNHVQD--KEKTVLLKK 713

Qy        358 KGSEMNASSIENVVIKSN---HLLSNNNNNKINRGKINDKI----ERNNILKNKSFDKPR 410
              |     : ::   : :|    |        :  :|    | |      |:||:||
Db        714 KKEINDKNTFSSCLINHNITTYTLQNGVNKNLNMLGIRDSIYKIDEKNNMLKE------- 766

Qy        411 EGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDN-----TINKKHNNNN------- 458
                        |:  || :  || ||| | ::    : |||      : :|:|:||
Db        767 --------CYNGNNDSNNKKKKKKKL-SFSCDIINDNITPYESDKEKNNSNNIKSMDIF 817

Qy        459 ----------NNNNDNNDNNNDNNNNNNNDSSSMV--NNMINHMINNNINN--------- 497
                        || : | |:  | :|  |::     :    | |   : |   | |
Db        818 NYVKRKSNLYNNLSSNRDSTVDMHNKYNSEEYINIQRTNKIYELSNKRIRNYKLYSMDEI 877

Qy        498 -NINVNN----NNNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLH 552
                :::       :| :||     :: | :    |     :     |    | |  ||  :   |
Db        878 FKVSLKEKKYIDNISNNMERVTYKNEM--INEKISKMDDILYPCDKNKSLNMSCPVI--- 932

Qy        553 KKNLRDNI-KLKNRSSCDNIMNKKKN---------NLHLARHSVGSKLTMFSYDPQKNK 601
                 : :|| : :|   :         |:||||                 : |: :            |:
Db        933 ---IENNISREENEKNSSVILNKKKNENMFNCVGRLHCHMGK-------------MNNQ 975

Qy        602 DNTFFKSNINKMEDNTPKDILYESRNVSN-MNGNVLLGLNKNTHHDI---STKDENHNDN 657
              || : : || | |:     |   | ||   |    |    :    :    :   :   | ||:  |
Db        976 DNIYDQGNIKKNEEEITKHDEYISREEKNKYNSKCIRNFDDYKYEQVLSYHTLDEDKKKN 1035
```

```
Qy        658 KIN--------------NGVINII------NNSN------------------------- 671
               :|               ||||| |      |||
Db       1036 DMNNLIDMNNEAIIETVNGVINNIILDRKDNNSRKDMEKEMEKEMEKKMEKEMEKVMEKE 1095

Qy        672 -------------VNSINNSNMNSINNSNMNSNSIYKSN--YNSNQSISDVQIRYVNE-- 714
                           | :||    | :||    |  :|||:|  |    |     | ::  |||
Db       1096 MEKVMEKEVEKELKNEMNNRMNNRMNNEMKNEINIYKNNEIYVDN----DKELEIVNEEK 1151

Qy        715 ---------------MDTS-NKNN-----DNIFFDAI--SC----DNNMYPNITNNNNNN 747
                             |: | | ||     :||  :  :  ||     :  |::  :  | |  :
Db       1152 KLIYPFNYESDVHKNMNMSININNCKDDYNNILKEYVDNSCLAQKEENIFRPLFNLNKKD 1211

Qy        748 N-----NNNNN-----------------NNNNIDVENYNNRD------------------ 767
                    |  ||                 | |   :  |:|  :
Db       1212 KVWKRFNIKNNIKTIIHNEEMKRIYQTINKNVFPIYNFNRYENFLINHLTYNFPKNDLFK 1271

Qy        768 -----GTNNSMKLYAYNSH--NLFQPDN---NKN-------TSNIQN------------- 797
                   ||   ||  |  |  | :    |    |:|        :|| |
Db       1272 LSYKVSMNNIRNLYIANKHINNNYDYMNKLYNQNIYTLKYQVANIDNDHHICKKGGGLDY 1331

Qy        798 INTN-----KNNQDGNVNYSMNFCHYN--------LNDK----NYLIDLNNKEQKDKN-- 838
              || |     || :|     |    ||          :||:   :|: |:  |   |:|
Db       1332 INMNISKECKNRKD--KTYLNKIFHYKKKKDARFFINDEIGSNDYMYDIKKKYSNDENNY 1389

Qy        839 ---------------------IHG-----CDNNIIQNR--------------NDFEKKK 857
                                   ||     |  |::: :               :||  ::
Db       1390 KLNEKMNISMSNDEDMIPTLNSEHGNNFPSCQPNLLEKKSTYIDLNLYDSNSMDDF-TEE 1448

Qy        858 KTNFYNNNNIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKD--------- 908
              | || || | :          |: | |:    : :    |: ||:  |   :|
Db       1449 KYNFVNNENDLF--------NTKRWKF----NFSKGKNLFNNKFFNVSNEDGVFSFFKNM 1496

Qy        909 NEMKDNHIKSNNNNSSSSSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNI 968
              |  :: :  :|:    | :||||  ||   ||   |     |     ||     : :   :|
Db       1497 NLFRELNKSNNSLKLESVKNSNNNCSNN-KGDDNIGNMENMNTTNVTIASDEHISTKGDI 1555

Qy        969 Y-------QNDDIIYTINSLNDYMSNTLLHFKEKYT----------YPTLSTNEDIY--- 1008
              :             || |:    |     :    |:   |: ::|               : :| |:
Db       1556 HDESFSRDDNDCILLKIEGRSKKYSDITLYNEDKSNLENDNETINEYENVCSNIDVNEWE 1615

Qy       1009 -----------NKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNVDNNNVDNNVDNN 1057
                         :||  |   :  : ::::   :|: ||  || |||  ||  |: || :|| ||
Db       1616 DKVNGTCNSVGDKETEKNNEKNNEKNNEKNNEKNNEKNNEKNNEKNN-EKNNEENNEGNN 1674

Qy       1058 DKNNVDNNNVDNDD-DDVDFHNIKNFNNNEYLSYFQKNVDTIINNCLNSLDISSMYDDTK 1116
              ::|| :||  :|:: :|:: ::||: |: :      ::|:   |: |  |::|: :  ||
Db       1675 EENNEENNEENNEENNDIEKNDIKDNNSGQ----VKENI--IVMNNTNNMDVDN--DDNN 1726

Qy       1117 EILNNI-------LLSKYKAE------KDNVIKKYINEDIKNMSL--------------E 1149
                 ||:          ::     |:|         ||::: |  |: |  |::                :
Db       1727 NNYNNVSTDEGIDIIKNIKSEMNDYIYNDNIMIKINNKSIDLMNIKNQKNEPFLNYTNEK 1786

Qy       1150 EIDKTAQSIY---------------EKRKVLLTKLLLLFKKNV---------DTQINNET 1185
              :|    :  | |                     ||    |   ||    |: :              || :  |  |
Db       1787 DIHMKSNSSYNVNDKMNLFNNNEKTEKNNTSLNDLLYKRKEELDDEKISEYKDTNLTNNT 1846

Qy       1186 SD---LRKDLVMCHICNNNPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLR---LLYQ 1239
                 :     | :|::         |:    :|     :     :     :     :  | :|:     :|:
Db       1847 FEHIAKRINLIL-----NDTIEFFQKHTYLHNGYGNVQVCKKNKRKLEKKKLKKWSCIYK 1901

Qy       1240 FLVVEYQNKSANSVLLNV-SSNNGDII---------LLNKKLVQDNIKNSMDHNNIHK 1287
                : :     | |: |: :     | | |              | |||         |    :|  |:::
Db       1902 --INKIVRKGAHGVVFSAWRSENVDFFNHSFFENLNLENKK------KGYIDETNVNE 1951
```

RESULT 37
AAB18315
ID AAB18315 standard; Protein; 2380 AA.
XX
AC AAB18315;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:173.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 395-401; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 2380 AA;

```
  Query Match             8.5%;  Score 579;  DB 21;  Length 2380;
  Best Local Similarity   21.1%;  Pred. No. 4.8e-24;
  Matches  302;  Conservative  208;  Mismatches  414;  Indels  510;  Gaps   67;

Qy        209 NDESSRSHAILNIDLKDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQTDGANINRSLL 268
                | ||:  |         |:||||    |  ||::  |      :: | |   :   :  | |   :
Db         20 NIESAEEH-------KNINKNIKNKK--FINIDNSNNCNNSNSNNSNSNNNNNNNNNIVR 70

Qy        269 ALKECIRAMDSDKNHIPFRDSELTKVLRDIFVGKSKSIMIANISPTISCCEQTLNTLRYS 328
                     | |  | ||  |         |::          |:| ::               :|   |
Db         71 NNNNFINA-DKKKNVILNEDDDI----------KNKELV----------DESFVNIFFYE 109

Qy        329 SRVKNFKNKSTCINEEDDTNTERISILDSK-GSEMNASSIENVVIKSNHLLSNNNNNKIN 387
               :  ||           |  | :| : |:|:: | | | | || :        :|: ::: :| |    |
Db        110 NYFKNL------FNLNDVSNNKVINIIEQKEGDERNADN----NLKNKNIVRDNINKIKN 159

Qy        388 RGKINDKIERNNILKNKSFDKPREGFTSTFGKY--SSLNDIDKIKKNKKKGLINYKSTLY 445
                   :|: :  ||.                    ||   : |||  ||         :   |     |:
Db        160 TRNVNEILIYNN-------------------KYIINFLNDTTKCKIEIANFISFYFFFLH 200
```

```
Qy        446 NDNTINKKHNNNNNNNNNDNNDNNNDNNNNNNNNDSSSMVNNMINHMIN-NNINNNINVNNN 504
                  : :||                     ||||     |    ||        :   :   |     |    |   :
Db        201 IKDILNK--------------NNDNGLMNKKKSS------LKDICNIKYIYKKIKTSKK 239

Qy        505 NNNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIF------------- 550
                ::|: :      |:|         |   ::  :      |    :|  |  ::
Db        240 YISSNDMDTCIRNYL-----YHIDKKNYPIIKKTKCPFLSNTKVLYNKRGYMASCPLTVK 294

Qy        551 ---LHKKNLRDNIKLK---NRSSCDNIMNKKKNNLHLARHSV-------GSKLTMFSYDP 597
                   || |:    ||||    |  |:    |   |  :|  ||:::  :     :       :    |
Db        295 GKIKHKTNISSKIKLKRERNDSNMFNNMIRKDNNMNVKQEQIHNNDTVNNNMTTNVDGCS 354

Qy        598 QKNKDNTF-------FKSNINKMEDNTPKDILYESRNVSNMN---------------GN 634
              :   ||||         ||    |  ::|     :       |:      :|    |                   |
Db        355 EPTHDNTFLNIEEEEFKMLKNYLKDVKERKKKYKKGYISTSNFISHGVRLGTTRSRIRGK 414

Qy        635 VLLGLNKNTH-----HDISTKDEN--HNDNKINNGVINIINNS-NVNSINNSNMNSINNS 686
                  ||  ||    |               :::  |   :       :   |::|   ||::    :|:     ::      |       |   |
Db        415 CLL-KNKKMHMYDDNEELNKKKKKKMNKDDRIENGIMEDVNDKRKLDCDNKIKFNDIEKE 473

Qy        687 NMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNNMYPNITNNNNN 746
              ::|                 |  |   :      |    :   ||  ||:|                  |||||
Db        474 DLN-------------ICDTE----NVDNNSNNNNN---------------NNNNNN 499

Qy        747 NNNNNNNNNNIDVENYNNRDGTNNSM--KLYAYN--------SHNLFQPDNN------- 789
               ||||||||||    :      |::           ||    ||           |: :   |||
Db        500 NNNNNNNNNNGYKKKIKNKNKNKNKKKNKLNNYNDNFVSVNGSYDNYSIDNNVINDEIR 559

Qy        790 ---KNTSNIQNINTNKNN---QDGNVNYSMNFCHYNLND-------KNYLIDLNNKEQKD 836
                   ||     ::  |    :|||     :|||    |    ::    :|:  :        |||:           :|:
Db        560 EKKKNNKEVK-IMVDKNNDTEKDGNKKYDTSY-SFNIKNTLSKVFYKNYVKRKGMIKQQH 617

Qy        837 KNIHGCDNNIIQNRNDFEKKKKTNFYNNNNIVIVNN--------NMGNNNSPR------- 881
                 |||     |    |   |  ::              |:     :|   :|::::                  |:|  :  ||
Db        618 NNIHNTHN--IHNTHNMH---NTHNIHNEKVVLLDDTKEKADPMNLGISFSPAGLLIPYH 672

Qy        882 -------MKYGLCGSHTS-------------------------IDNMKNNEMK------ 902
                         ::   :     |||                                |:|:  :|   |
Db        673 LGVSSLLIEKNILNMHTSIAGSSAGSICACCLSVGLSVNKCYFLIENIISNVYKHGCYQK 732

Qy        903 -----NNEMKDNEMKDNHIKSNN----------------------NN--------SSSSSS 928
                    | |:      :|::|    ||                                 ||              |:    :|
Db        733 LENILNIELNKYLYEDSYIYLNNRIGNVFVGITQILPYYKKLNINNFYDDNDLISAIIAS 792

Qy        929 SNNNIYNNINDDDTFQNDYCHN------------DNTFT---------------IRRKNN 961
               |   :|  :  |       |:|    |   :                            |||                    |    |||
Db        793 CNIPMYLSSNIFVNFRNKKCIDGFFSTKKKDFGCPNTRTERIIKVSPFDSDYVGIGNKNN 852

Qy        962 TNIN-----------------------SNIYQNDDIIYTINSLNDYMSNTLLHF----K 993
              :  |:                                        :|::       |    ::  |   :|   :|   |:     :      :        |
Db        853 SVISPHLIKYNHILFLFICVKNIFHKYINNLWIEKDYLFLIENLKDILERKIFDYYTFVK 912

Qy        994 EKYTYPTLSTNEDIYNK------EMEGKHIRLDDQDKYDD------NDNNNVDNNNKNNV 1041
               :|:   |    ||  |   :|            |  ||:      :|  |:  :|           |::|    |     :||
Db        913 RYFTF--LRKNETIDDKYEEEEYEDEGEEYEEEDDDEEEDEEEYGHNNDNQDDEGDKNKT 970

Qy       1042 DNNVD------NNNVDNNVDNND--KNNVDNNNVDN----DDDD-------VDFHNI--- 1079
                |    :             |||  :||:  ||:       ||:  |||:::             :  |                  | ||
Db        971 TNEKNKKKKNKNNNNNNNIFNNNIFNNNIFNNNMNSCVGVSEKDFISTSIVASFANIKRQ 1030

Qy       1080 ----------------------------KNFNNNEYLS--------------------- 1089
                                                  ||  |  |  |::
Db       1031 MNEKIEKRKNEKKEKKEKLQRKNMNKCSKNRNRNRYINKDSNIHLMNLIRIKFKNLNYMN 1090

Qy       1090 --------YFQKNVDTIINNCLNSLDISSMYDDTKEILNNILLSKYKAEK------DNVI 1135
                              |  :  |   |        :: ::  :  |:  :    ::|   ::|||   :|                        ::
Db       1091 MNSFEIELYLKINNDIFLQFNKHNYNVQNFYNFSITLIN--IMSKYYSENFYAYNLEKIV 1148

Qy       1136 KKYINEDIKNMSLEEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMC 1195
                 |::     :  |   :  |  |:|        |   :  :                    :|   :|     |:: |    |  ::
Db       1149 YKFL---LNNKNFEYIEKQ----YSSKEDM---------NELDILVN--TYDMKYDKIIE 1190

Qy       1196 HICNNN--PDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFL----------- 1241
                    : ||         |::    |   |  :|:  |||  |           ::|:          |::   :||
Db       1191 FLKNNGYLKIDRY-IYFYPKLKTDII-LFFFKEIFLNDNILKIDRKFLKKNITIMIEVLK 1248

Qy       1242 ---VVEYQNKSANSVL-LNVSSNNGDIILLNKKLVQDNIKNSM-------DHNN 1284
                             ||     :          |:            |          |   :  :||         :         |:                ||||
Db       1249 EIFFKEYVKRCITKVIFFPVHMKEHDHV-MNKNYYNNQYVNNSNMFNTRGDHNN 1301
```

RESULT 38
AAB18272
ID   AAB18272 standard; Protein; 2500 AA.
XX
AC   AAB18272;
XX
DT   07-NOV-2000 (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:129.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 302-309; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   2500 AA;

Query Match   8.4%; Score 577.5; DB 21; Length 2500;
Best Local Similarity 24.1%; Pred. No. 6.1e-24;

```
Matches  248;  Conservative  164;  Mismatches  389;  Indels  229;  Gaps   52;

Qy        355 LDSKGSEMNASSIENVVIKSNHLLSNNNNNKINRGKI----NDKIERNNILKNKSFDKPR 410
              :|  |   | :|: |:   :|:: :|  |    ::| |    |:  ::|||      |
Db         33 IDEPGVLSNMNSVSNISTSTNNIGTNTMNFNNSKGFIINPFNENYKKNNIC--TYLDHES 90

Qy        411 EGFTSTFGKYSSLNDIDKIKKNKKKGLINYKSTLYNDNTINK-------KHNNNNNNNND 463
                    :|   | :|:: :  :   :|  : :   | :|:              |  |  |
Db         91 TNINGGVNQYD--NHMDQMNQMNQTNQMNQMNQMNQTNQMNQMNQMNQMNQTNQMNQTNQ 148

Qy        464 NNDNNNDNNNNNNNDSSSMVNNMINHMINN--------NINNNINVNNNNNNNNNNN-- 512
                 |  |  |  |  |  |  |   |:                    | :  ::  |||  :| |
Db        149 MNQTNQMNQTNQMNQTNQMNQMNIQHQRNSVNAPNIYIQNFDQNCDIYYNNNGKSNGNLN 208

Qy        513 --NSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNRSSCDN 570
                 |  |  |  |          |  |  |   |   |   |:||         ||  :  ||:
Db        209 VQQSDNAHNPL-------IYDISEL--YNREKNEEQKTIF------RD--EYSNRTIIKA 251

Qy        571 IMNKKKNNLHLARHSVGSKLTMFSYDPQKN---------KDNTFFKSNINKME------ 614
              ::||  |    :      :  |  ||:  :|                 | :    | |: :
Db        252 LINKITNTPMINNSVKNIEDTNSSYNTDENVYNVCSMDEYTTNKYISKNYNENDQVIVQG 311

Qy        615 DNT----PKDILYESRNVS----NMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINI 666
              :||      : :|:  |:|     ::  |:    :  | :||         : :    |:   | :
Db        312 NNTVPENDNNEIYKKENLSIFQDSLKDNI---VEYNAYHD------SRHHKPIDEQVAHY 362

Qy        667 INNSNVNSINNS-NMNSINN-----SNMNSNSIYKSNYNSNQSISDVQIRYV-NEMDTSN 719
              |||  |:  |:  |  |||  ||          :|:|  ||  :  :  |   |:   :   |  |:  :  |
Db        363 INNYYTNNNNDPYNRNSTNNNGIAENNINVNSAF-NQYKENKQYYDLLNTFTGNIMERKN 421

Qy        720 KNNDNIFFDAISCDNNMYPNITNNNNNNNNNNNNNNNNID--------VENYNNRD---- 767
                |: ::     |::   ||   :||    |:     :|||::            :  |:|  ::
Db        422 IMMQNVDYNERINGNSI--NIQGSNNQQMNDQLVDNNNVNMCLMQGPYINNHNMKNFYMG 479

Qy        768 ---GTNNSMKLYAYNSHNLFQPD--------NNKNTSNI--QNINTNK--NNQDGNVNY- 811
                   | ||:|      ::|:   :              ||  |:|:   ||  ||   ||  ||  |
Db        480 SNYGGNNNMVHNIMGTNNMVHNNMVHNNMGTNNMGTNNMGNNNIGTNNMGNNNMGNNNIG 539

Qy        812 SMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNN---NIV 868
              : |  |  | |: :  |  |  ||    :    :    || : |              |:  ||   |
Db        540 TNNMVHNNMGNNN--IGTNNMVHNNMGNNYMGNNYMGN----------NYMGNNYMGNNY 587

Qy        869 IVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNN---NNSSS 925
              :|:|||  ||         :    ::      |||:||  |   |   |   |   |   |:: :||     ||   :
Db        588 MVHNNMSTNN-------MGNNYMGNDNMRNNNMGTNNMGTNNMGTNNMGTNNMGTNNMGT 640

Qy        926 SSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDII-YTINSLNDY 984
              ::       ||: ||          |:|  |||              :||   |         | ||  |:  :|:  :
Db        641 NNMGTNNMGNN-----NMGNNYIGNDN-----MRNNHII-------DYIINYMVNNMVNN 683

Qy        985 MSNTLLH--FKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNK-NNV 1041
               | ||:::             |       :    :: |                      |   ||: |:|  |::
Db        684 MVNTMVNNMVNNMVNYMVNNMVNNMVN-----------------NMGNNMGNHNMINHM 725

Qy       1042 DNN-VDNNNVDNNVDNND--KNNVDNNNVDNDDDDVDFHNIKNFNNNEYLSYFQKNVDTI 1098
                 |: :  |  |:  ||:::|:    ||   ||    |:::||:     :|  |  |  |   |          :
Db        726 GNDRIGNYNMGNNLNSNNYMVNNYSNNTYGNNNNDVN----RNMNYN-YAGYNHMYARCL 780

Qy       1099 INNCLNSL---------DISSMYDDTKEILNNILLSKYKAEKDNVIKKYINEDIKNMSL 1148
                 || :|:                  ||:::  |||     |:|:    |  :|   :  :  :   :: |  |:
Db        781 NNNNMNNTPQYIIPDNKNKISAVHPFTKETNTGIILN--NASQDYTLSRSLGYNL-NFSM 837

Qy       1149 ----EEIDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLR--KDLVMCHICNN-- 1200
                       :  :  |   :|            |            :|  |  |         |    :|    :
Db        838 IQSENDFNSTPSNIEPVNNQPLNVEPAIFE-----AVNLELFDAESLNDQYVCDENSNTD 892

Qy       1201 ---------NPDDQFHFYAYSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSAN 1251
                                 ||  |   |  |:   |       |:|                ||   |         :    ||  :   |
Db        893 VIKSKPLVDNPLDDEHIYS-EHLNNRILNDETLPSAQLNVETL-----YGEHEYNEQRMN 946

Qy       1252 SVLLNVSSNN 1261
                 :|   |
Db        947 EQRINEQRMN 956
```

```
RESULT 39
AAW03625
ID   AAW03625 standard; peptide; 342 AA.
XX
AC   AAW03625;
XX
DT   25-MAR-2003  (updated)
DT   17-DEC-1996  (first entry)
XX
DE   Human lutropin-choriogonadotropin GPR N-terminal sequence.
XX
KW   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW   muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW   odorant; cytomegalovirus; serotonergic.
XX
OS   Homo sapiens.
XX
PN   US5508384-A.
XX
PD   16-APR-1996.
XX
PF   09-SEP-1993;   93US-0118270.
XX
PR   09-SEP-1993;   93US-0118270.
PR   10-SEP-1992;   92US-0943236.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Murphy RB,  Schuster DI;
XX
DR   WPI; 1996-208785/21.
XX
PT   New dopamine receptor peptide - useful as antipsychotic agent, e.g.
PT   for treating schizophrenia
XX
PS   Disclosure; Fig 8B(2); 184pp; English.
XX
CC   Proteins AAW02657-W02720 represent a range of G-protein coupled receptor
CC   (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other GPR proteins.  The peptides
CC   AAW03578-W03651 represent the N-terminal fragments of the above
CC   proteins. The receptor proteins were used to design polypeptides, pref.
CC   based on the transmembrane domains, for use in G-protein coupled receptor
CC   ligand binding assays.  The polypeptide fragments retain biological
CC   activity such as binding a GPR ligand or modulating GPR ligand binding to
CC   a GPR (see AAW02747-W02999 for examples of polypeptide fragments).
CC   The polypeptide fragments can be used in compositions for treating
CC   subjects suffering from a pathology related to a GPR abnormality e.g. a
CC   psychotic disorder such as schizophrenia.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   342 AA;

  Query Match             8.4%;  Score 576.5;  DB 17;  Length 342;
  Best Local Similarity   24.9%;  Pred. No. 7.4e-25;
  Matches  159;  Conservative   43;  Mismatches   63;  Indels  373;  Gaps   14;

Qy        446 NDNTINKKHNNNNNNNNDNNDNNNDNNNNNNNDSSSMVNNMINHMINNNINNNINVNNNN 505
              |:|  |   :|||||||:||:|||:||||||             ||| |||  | ||||
Db         34 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-------------NNNNNNNNNNNN 79

Qy        506 NNNNNNNNSHNNHLPQPNYAFTDTSDFSSLDDMNCHLNNNDKSIFLHKKNLRDNIKLKNR 565
              ||||||||::||                        |||
Db         80 NNNNNNNNNNN------------------------NNN------------------- 94

Qy        566 SSCDNIMNKKKNNLHLARHSVGSKLTMFSYDPQKNKDNTFFKSNINKMEDNTPKDILYES 625

Db         95 ------------------------------------------------------------ 94
```

```
Qy        626 RNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKINNGVINIINNSNVNSINNSNMNSINN 685
                                          |:|:|  ||       ||:| |: ||:| |: ||
Db         95 --------------------------NNNNNNNNN------NNNNNNNNNNNNNNNNNN 121

Qy        686 SNMNSNSIYKSNYNSNQSISDVQIRYVNEMDTSNKNNDNIFFDAISCDNMMYPNITNNNN 745
              :| |:|:   :| |:|                 | ||:|                |||
Db        122 NNNNNNNNNNNNNNNN-----------------NNNNNN-----------------NNNN 147

Qy        746 NNNNNNNNNNNNIDVENYNNRDGTNNSMKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQ 805
              |||||||||||    | || :  ||               || | :|  | | | ||
Db        148 NNNNNNNNNNN----NNNNNNNNNN---------------NNNNNNNNNNNNNNNNNN 187

Qy        806 DGNVNYSMNFCHYNLNDKNYLIDLNNKEQKDKNIHGCDNNIIQNRNDFEKKKKTNFYNNN 865
              : | |                                                   |||
Db        188 NNNNN---------------------------------------------------NNN 195

Qy        866 NIVIVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNNNNSSS 925
              |    |||  |||                 |  ||   ||            ||||:::
Db        196 N----NNNNNNNN---------------NNNNNNNNNN-------------NNNNNNN 221

Qy        926 SSSSNNNIYNNINDDDTFQNDYCHNDNTFTIRRKNNTNINSNIYQNDDIIYTINSLNDYM 985
              ::::|||  || |                        || | |:|
Db        222 NNNNNNNNNNNNN--------------------NNNNNNNN----------------- 242

Qy        986 SNTLLHFKEKYTYPTLSTNEDIYNKEMEGKHIRLDDQDKYDDNDNNNVDNNNKNNVDNNV 1045
                                                       |:||| :||| || :||
Db        243 ------------------------------------------NNNNNNNNNNNNNNNNNN 260

Qy       1046 DNNNVDNNVDNNDKNNVDNNNVDNDDDDVDFHNIKNFN 1083
              :||| :|| :||: || :|||:   : :      :||:
Db        261 NNNNNNNNNNNNNNNNNNNNNLPTKELNFSHSISENFS 298
```

RESULT 40
AAB18218
ID   AAB18218 standard; Protein; 2010 AA.
XX
AC   AAB18218;
XX
DT   07-NOV-2000 (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:75.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999; 99WO-US26796.
XX
PR   05-NOV-1998; 98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 177-182; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 2010 AA;

```
  Query Match             8.3%;  Score 569.5;  DB 21;  Length 2010;
  Best Local Similarity   21.3%;  Pred. No. 1.4e-23;
  Matches  287;  Conservative  192;  Mismatches  404;  Indels  467;  Gaps   59;

Qy        224 KDINKNTSLGKIAFIDLAGSERGADTVSQNKQTQ---TDGANINRSLLALKECIRAMDSD 280
              |:: |       |          ::| |   | ||: :     :   |:| |       | : ||:
Db          8 KEVEKKEEKEK-----KGAKDKGKDLFSLNKKRERKKKESQKIDRYL------INSCDSN 56

Qy        281 KNHIP----------------------FRDSELTKVLRDIFVGKSKSIMIANISP---- 313
              |::                        |     :|||:     ::    | || |
Db         57 KSNYSCCYLNNECFVKNISICKKCMFSYFEFKNVTKVIYMRHGARTPKKKIKNIWPFKEG 116

Qy        314 ----TISCCEQTLNTLRYSSRVKNFKNKSTCINEEDDTNTE--RISILDSKGSEMNASSI 367
                  |    :|::     |   :      ||    :|:: :      ||:   :     : |    :
Db        117 KGDLTFLGFQQSIKVGEYLRKYYYTFNK---LNKKYNKRERGLRINNKEKGYIKKNKCDV 173

Qy        368 ENVVIKSNHLLSNNNNNKINRGKINDKI---ERNNILKNKSFDKPREGFTSTFGKYSSLN 424
              :       :  :|||||  |   ||:|    :|: :|| ::|    :|::  :     :| |
Db        174 KKCKTLYKNKYNNNNNNNNNNYVINEKYNGSNKNDYVKNNTYD--NKGYSYLYDLSTSFN 231

Qy        425 D---------------------------------------------------------- 425
              :
Db        232 ELENRKRKLHKFPYLRDFIYYEKYFLKINKRSNKHQRKVFIKIKRRRRNNILKIWIHQHL 291

Qy        426 IDKIKKNKKKGLINYKSTLYNDNTINKK-----------HNNNNNNNNDNNDNNNDNNN 473
              |:|:|| |  | : ||               :: |          ||:::||:|||:||||:||
Db        292 INKMKKIKNKNMNNYNKCYIKFSSIRKRGYHKMENIECNNKNNDDDNNDDNNNNNNDDNN 351

Qy        474 NNNNDSSSMVNNMINHMINNNINNNINVNNNNNNNNNNNSHNNH-----------LPQP 522
              |||||  :                ||:  |||  |  |||::|||||:   ||:
Db        352 NNNNDDN-----------NNDDNNNDNNNNNDDNNNNNDDDNNYYYYNYNNDETPFNNK 400

Qy        523 NYAFTDTSDFSSL-------DDMNCHLNNNDKSIF--------LHKKNLRDNIKLKNRSS 567
              :: :  |    ::             |  | :  ||   | :|          ::||  |   | |: |  :
Db        401 SFNYADMLKYTKYYYKNILKDKKNIYTNNKKKELFFPLMEHLYMYKKKLLIN-KMKEK-- 457

Qy        568 CDNIMNKKKN--------NLHLHARHSVGS---KLT---------------MFSYDPQKN 600
              ||  |||             | :|   :   |   |||                   | :
Db        458 --NIKKKKKKYDKIIKLINKYLCIKTTNSERCKLTAYGIICGILGISEYIYFFFF----- 510

Qy        601 KDNTFFKSNINKMEDNTPKDILYESRNVSNMNGNVLLGLNKNTHHDISTKDENHNDNKIN 660
              |||||  :|   ||                :|          |     :  ||::  |   | |
Db        511 --ILFFKSNYDKTNDNNIDTYTKRKEKKKCLNKRSKCFQNWILNRDITSGQYNCID-KNT 567

Qy        661 NGVINIINNSNVNSINNSNMN------------SINNSNMNSNSIYKSN-YNSNQSISD 706
              | |  |    |:     |    |                |   : ||   :::||      ||
Db        568 APVKNYIIGENLCGENGCGKNGCGDILRGDILCGDILRGDNNSIPLFRSNRIFCKQS--- 624

Qy        707 VQIRYVNEMDTSNKNNDNIFFDAI-----SCDNNMYPNITNNNNNNNNNNNNNNNIDVE 761
              :| : :|:            |:|: |       | | :||                   | :|:
Db        625 -KITFCDEL--------YIYFNKILKRLQSLD-DMY-----------------KINHEVK 657
```

```
Qy        762 NY-NNRDGTNNS-MKLYAYNSHNLFQPDNNKNTSNIQNINTNKNNQDGNVNYSMNFCHYN 819
              : |::|  |||  | |  | :  : | || ::: ::                    ||
Db        658 MFGNDKDVLNNSYKKCYDKNDYGSY-PSYNKYSNDYKS------------------HYV 697

Qy        820 LNDKNYLIDLNNKEQKDKNIHG--CDNN--IIQNRNDFEKKKK-------TNFYNNNNIV 868
              :              :| ||:    | |   |:: | : |||||          | ||||::
Db        698 I-------------KKMKNVKSVQCSNESIILKERQENEKKKKKKKKMENTFINNNNLM 744

Qy        869 --------IVNNNMGNNNSPRMKYGLCGSHTSIDNMKNNEMKNNEMKDNEMKDNHIKSNN 920
                      :: |  ||                  :|:|     || :
Db        745 YNINVFFDLIINERGN-----------FQFFYNNIKKKRQKNEK-------------- 777

Qy        921 NNSSSSSSSNNNIYNNINDDDTFQ-------NDYCHNDNTFTIRRKNNTNINSNIYQNDD 973
                           |:||       ||       |::       |     ||  ||::
Db        778 ------GLEEWNVYN------IFQLYMKYILNEFSKFFKLFKFLNKNVENIDN------- 818

Qy        974 IIYTINSL----NDYMSNTLLHFKEKYTYPTLSTNEDIYNK-EMEGKHIRLDDQDKYDDN 1028
                 | ||:     |  |   | ::| |: :    : :.| :   |   :  | |   : :::  ||
Db        819 ---TFNSITNIYNKYYINMVVHRKDCFEKKQIHSKEHMMKKIHLRDKFIEYEKENEIIDN 875

Qy       1029 DNN-NVDNNNK---NNVDNNVDNNNVDNNVDNND------------------------K 1059
               || |:||  |    || :| :|||:: :: ||                          |
Db        876 CNNINMDNKKKEINNNYNNMIDNNNIEIDMSNNFIFTYYYIFYLLNYMDTYIQFLFYYLK 935

Qy       1060 NN---------VDNNNVDNDDDDVDFHNIKNFNNN--------EYLSYFQKNVDTII--- 1099
              |             : |::           | ||   |:           : |: : |:   |:
Db        936 NTYILFSVVKVAERNSLMLKTLKTKNHYIKKLRNHIIHNSDVYKILNNYYKDEIFIVYDI 995

Qy       1100 ----NNCLNSLDISSMYDDTK-----EILNNILLSKYKAEKDNVIKKYINEDIKNMSLEE 1150
                  ||:|: ||  :|:| |        : | || :      :|:              : |  :
Db        996 TKWTENCMNTTDI--LYNDVKKNTKIDDLENIDIPIITNDKEEY-------HVNNSIISV 1046

Qy       1151 IDKTAQSIYEKRKVLLTKLLLLFKKNVDTQINNETSDLRKDLVMCHICNNNPDDQFHFYA 1210
              :  |     |:|: :| |   ::|                 ||:|  : |:  |  |       ::
Db       1047 LKKHNSSVYKLKKKLKNSIIL--------------KDLKK--LNCNFINKNYIHNTNYDK 1090

Qy       1211 YSRLEKDIINLIMLRQIWCESENLRLLYQFLVVEYQNKSANSVLLNVSSNN--------- 1261
              :::: :| |           : :|:::: :| :   |      : | ||:: |
Db       1091 HNKIYQDKIKNWTYHPFHNKKKNVKIIKKF-ISAYDAYIYHGVNLNLNFNRAYEKLSQHP 1149

Qy       1262 -GDIILLNKKLVQDN------IKNSMDHNN 1284
                 | |: |:   |:|        ||   : ||
Db       1150 PSSIDLIKKEYGQNNYIINGEIKKYEEQNN 1179
```

Search completed: October 2, 2003, 16:17:06
Job time : 105 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 2, 2003, 16:28:56 ; Search time 52 Seconds
(without alignments)
2382.024 Million cell updates/sec

Title: US-10-006-780-2
Perfect score: 1288
Sequence: 1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	17	1.3	391	2	S14577	asparagine-rich pr
2	17	1.3	419	2	S14508	asparagine-rich pr
3	17	1.3	666	2	S50452	hypothetical prote
4	17	1.3	749	2	S23467	probable long-chai
5	17	1.3	758	2	S45825	cell division cont
6	17	1.3	943	2	S68824	rngB protein, cyto
7	17	1.3	1093	2	T18275	1-phosphatidylinos
8	17	1.3	1139	2	S71092	UV-damaged DNA bin
9	17	1.3	1178	2	S57698	regulatory protein
10	17	1.3	1338	2	T18287	protein-tyrosine k
11	17	1.3	1457	2	T14577	protein kinase Yak
12	17	1.3	1671	2	S71628	sensory transducti
13	17	1.3	1969	2	T08875	histidine kinase h
14	17	1.3	4550	2	T18440	hypothetical prote
15	16	1.2	188	2	T02389	hypothetical prote
16	16	1.2	317	2	S05356	hypothetical prote
17	16	1.2	328	2	T02046	prolamin box bindi
18	16	1.2	373	2	T52593	squamosa promoter
19	16	1.2	401	2	T18285	hypothetical prote
20	16	1.2	490	2	A46391	cAMP receptor subt
21	16	1.2	490	2	S52830	HMS1 protein - yea
22	16	1.2	495	1	S52641	heat shock transcr
23	16	1.2	566	2	S53813	RNA helicase - sli
24	16	1.2	580	2	T18439	hypothetical prote
25	16	1.2	600	2	T18446	hypothetical prote
26	16	1.2	686	2	A71607	Mtn3/RAG1IP-like p
27	16	1.2	720	2	T51007	hypothetical prote
28	16	1.2	758	2	S54522	hypothetical prote
29	16	1.2	831	2	T08611	hypothetical prote
30	16	1.2	947	2	T08605	hypothetical prote
31	16	1.2	961	2	S67568	probable membrane
32	16	1.2	962	2	JC5808	G protein-coupled
33	16	1.2	964	2	S48404	probable membrane
34	16	1.2	1075	2	S54067	probable membrane
35	16	1.2	1225	2	T09057	probable protein-h
36	16	1.2	1256	2	S14556	asparagine-rich pr
37	16	1.2	1271	2	T08607	hypothetical prote
38	16	1.2	1336	2	T18288	ABC transport prot
39	16	1.2	1365	2	S14871	suppressor two of
40	16	1.2	1390	2	T14004	trfA protein - sli
41	16	1.2	1402	2	T17456	cell surface prote
42	16	1.2	1584	2	T18276	protein-tyrosine k
43	16	1.2	1585	2	T18274	1-phosphatidylinos
44	16	1.2	1619	2	T18499	hypothetical prote
45	16	1.2	1743	2	T18279	multidrug resistan
46	16	1.2	1817	2	D71606	hypothetical prote
47	16	1.2	1844	2	D71612	hypothetical prote
48	16	1.2	1855	2	S41649	DNA polymerase - m
49	16	1.2	1858	2	T18273	1-phosphatidylinos
50	16	1.2	1905	2	T18267	multidrug resistan
51	16	1.2	2150	2	S71629	sensory transducti
52	16	1.2	2380	2	E71604	hypothetical prote
53	16	1.2	2523	2	T18477	hypothetical prote
54	16	1.2	3844	2	T18402	asparagine/asparta
55	16	1.2	3848	2	T17414	TipC protein - sli
56	16	1.2	4981	2	T18489	hypothetical prote
57	15	1.2	125	2	E43692	T3C protein - rabb
58	15	1.2	451	2	A23535	clustered asparagi
59	15	1.2	612	2	S65196	probable membrane
60	15	1.2	837	2	S36713	CCR4 protein - yea
61	15	1.2	888	2	S64016	probable regulator
62	15	1.2	1655	2	T13998	gene mastermind pr
63	15	1.2	2278	1	S56274	FAB1 protein - yea
64	15	1.2	2539	2	B71619	hypothetical prote
65	15	1.2	3394	2	T18501	hypothetical prote
66	14	1.1	150	2	D48835	kinesin-like prote
67	14	1.1	247	2	T34214	hypothetical prote
68	14	1.1	264	2	A44969	circumsporozoite p
69	14	1.1	375	2	T00929	squamosa-promoter
70	14	1.1	668	2	T02791	mitotic centromere
71	14	1.1	700	1	S09748	kinesin-related pr
72	14	1.1	744	2	T06048	kinesin-related pr
73	14	1.1	754	2	S48020	kinesin-related pr
74	14	1.1	777	2	C85065	kinesin-like prote
75	14	1.1	793	2	S34830	kinesin-related pr
76	14	1.1	832	2	T38749	kinesin-like prote
77	14	1.1	858	1	A42239	adenylate cyclase
78	14	1.1	913	2	T18503	hypothetical prote
79	14	1.1	955	2	A47334	LcKin kinesin-rela
80	14	1.1	987	2	B96766	protein kinesin F2
81	14	1.1	1056	2	C96661	kinesin-like prote
82	14	1.1	1245	2	D71613	GAF domain protein
83	14	1.1	1662	1	H71402	probable kinesin -
84	14	1.1	1770	2	S56221	hypothetical prote
85	14	1.1	1891	2	T13594	hypothetical prote
86	14	1.1	2010	2	B71616	phosphatase (acid
87	13	1.0	201	2	S54475	hypothetical prote
88	13	1.0	367	2	T46021	hypothetical prote
89	13	1.0	426	2	T04318	homeobox protein L
90	13	1.0	554	2	T50118	kinesin-related pr
91	13	1.0	608	2	T18437	hypothetical prote
92	13	1.0	669	2	S64795	suppressor protein
93	13	1.0	697	2	T16306	hypothetical prote
94	13	1.0	708	2	A53185	G-box-binding fact
95	13	1.0	728	2	S21913	BRcore-TNT1-Q1-Z1
96	13	1.0	851	2	S67285	NUD1 protein - yea
97	13	1.0	904	2	S69691	hypothetical prote
98	13	1.0	1065	2	T13230	dachshund isoform
99	13	1.0	1072	2	T13232	dachshund protein
100	13	1.0	1074	2	T13229	dachshund protein
101	13	1.0	1075	2	T45570	kinesin-like prote
102	13	1.0	1081	2	T13231	dachshund protein

103	13	1.0	1083	2	T26844	hypothetical prote
104	13	1.0	1151	2	A41529	GRR1 protein - yea
105	13	1.0	1335	2	T18289	racGAP protein - s
106	13	1.0	1920	2	T13893	gene hindsight pro
107	13	1.0	2269	2	T18472	hypothetical prote
108	13	1.0	2573	2	D71614	hypothetical prote
109	12	0.9	78	2	S23689	erythrocyte membra
110	12	0.9	212	2	S49807	hypothetical prote
111	12	0.9	239	2	S14470	asparagine-rich pr
112	12	0.9	431	2	S50977	hypothetical prote
113	12	0.9	537	2	A23770	asparagine-rich pr
114	12	0.9	590	2	A26638	homeotic protein D
115	12	0.9	603	2	F96797	probable GT-like t
116	12	0.9	647	2	S06450	steroid hormone re
117	12	0.9	671	2	T10755	kinesin-related pr
118	12	0.9	716	1	A44259	kinesin-related pr
119	12	0.9	747	2	T23607	hypothetical prote
120	12	0.9	764	2	H71607	hypothetical prote
121	12	0.9	813	2	T46242	kinesin-like prote
122	12	0.9	914	2	S46593	finger protein AZF
123	12	0.9	1341	2	S50366	probable membrane
124	12	0.9	1516	2	E71619	RAD2 endonuclease
125	12	0.9	1570	2	T18272	1-phosphatidylinos
126	12	0.9	1864	2	T18485	hypothetical prote
127	12	0.9	2452	1	RNZQ2L	DNA-directed RNA p
128	12	0.9	2910	2	T28156	DNA-directed RNA p
129	12	0.9	3724	2	T18427	hypothetical prote
130	11	0.9	95	2	F41298	kinesin-like prote
131	11	0.9	234	2	S14469	asparagine-rich pr
132	11	0.9	280	2	C86317	protein T10O22.23
133	11	0.9	330	2	B48835	kinesin-like prote
134	11	0.9	331	2	G96785	protein F10A5.24 [
135	11	0.9	443	2	A54813	cAMP receptor CAR4
136	11	0.9	448	2	S05355	hypothetical prote
137	11	0.9	461	2	T10265	arabinogalactan-pr
138	11	0.9	472	2	A26357	homeotic protein C
139	11	0.9	519	2	I54523	kinesin-related pr
140	11	0.9	522	2	T18504	hypothetical prote
141	11	0.9	549	2	T20720	hypothetical prote
142	11	0.9	601	2	T00119	probable transcrip
143	11	0.9	622	2	A57281	kinesin-like motor
144	11	0.9	670	2	T29898	kinesin protein OS
145	11	0.9	672	2	S54351	kinesin osm-3 - Ca
146	11	0.9	676	2	S61977	transcription fact
147	11	0.9	690	2	T34149	hypothetical prote
148	11	0.9	703	2	A86319	F25I16.11 protein
149	11	0.9	722	2	D71607	VPS45-like protein
150	11	0.9	729	1	A34796	kinesin-related pr
151	11	0.9	784	2	T40594	probable kinesin-l
152	11	0.9	805	2	S64238	kinesin-related pr
153	11	0.9	817	2	T50240	kinesin-like prote
154	11	0.9	836	2	T18460	hypothetical prote
155	11	0.9	861	2	T00434	probable kinesin h
156	11	0.9	920	2	T08853	protein phosphatas
157	11	0.9	921	2	T01775	hypothetical prote
158	11	0.9	932	2	T30099	hypothetical prote
159	11	0.9	935	2	T51930	kinesin [imported]
160	11	0.9	968	2	T51933	kinesin motor prot
161	11	0.9	975	2	T08606	protein phosphatas
162	11	0.9	987	2	T51360	kinesin-like heavy
163	11	0.9	1025	2	T18376	multidrug resistan
164	11	0.9	1032	2	B86224	hypothetical prote
165	11	0.9	1122	2	F71606	probable multiple
166	11	0.9	1225	2	A56514	chromokinesin - ch
167	11	0.9	1226	2	I51617	kinesin-like prote
168	11	0.9	1229	2	T48959	kinesin-like prote
169	11	0.9	1230	2	T18256	probable serine/th
170	11	0.9	1230	2	T18259	serine/threonine p
171	11	0.9	1231	2	A54803	microtubule-associ
172	11	0.9	1265	2	T03792	kinesin-related pr
173	11	0.9	1265	2	T07397	kinesin heavy chai
174	11	0.9	1269	2	T14476	pslA protein - sli
175	11	0.9	1351	2	C71607	hypothetical prote
176	11	0.9	1388	2	T30335	KLP2 protein - Afr
177	11	0.9	1459	2	T30196	kinesin motor prot
178	11	0.9	1979	2	C71622	hypothetical prote
179	11	0.9	1997	2	F71607	DNA helicase II BR
180	11	0.9	2061	2	T13751	transcription fact
181	11	0.9	2339	2	A45597	DNA-directed RNA p
182	11	0.9	2457	2	T18492	hypothetical prote
183	11	0.9	2708	2	T09079	probable chloroqui
184	11	0.9	3973	2	B71612	hypothetical prote
185	10	0.8	97	2	B96580	F15I1.5 [imported]
186	10	0.8	100	2	S23686	erythrocyte membra
187	10	0.8	240	2	AI1305	metal cations ABC
188	10	0.8	240	2	AI1677	metal cations ABC
189	10	0.8	272	2	S23684	erythrocyte membra
190	10	0.8	368	2	S46601	hypothetical prote
191	10	0.8	390	1	B49070	ecdysone-inducible
192	10	0.8	398	1	R3BYM1	ribosomal protein
193	10	0.8	419	2	T18420	hypothetical prote
194	10	0.8	522	2	B45268	interleukin-9 rece
195	10	0.8	575	1	JH0827	glutamate decarbox
196	10	0.8	587	2	T19893	hypothetical prote
197	10	0.8	587	2	T29324	hypothetical prote
198	10	0.8	613	2	T15489	hypothetical prote
199	10	0.8	623	1	S33167	gene pointed prote
200	10	0.8	633	2	T04179	hypothetical prote
201	10	0.8	648	1	JQ1150	protein kinase (EC
202	10	0.8	695	2	T24950	hypothetical prote
203	10	0.8	716	2	T26998	hypothetical prote
204	10	0.8	718	1	S33168	gene pointed prote
205	10	0.8	720	2	JQ1676	ABI3 protein - Ara
206	10	0.8	726	2	S67044	NFI1 protein - yea
207	10	0.8	786	2	T18469	hypothetical prote
208	10	0.8	823	2	S48986	probable protein k
209	10	0.8	864	2	A49070	ecdysone-inducible
210	10	0.8	928	2	T16759	hypothetical prote
211	10	0.8	974	1	A40213	optic lobe develop
212	10	0.8	1068	2	S64015	pleiotropic drug r
213	10	0.8	1121	2	T13750	iinesin-like prote
214	10	0.8	1121	2	T13796	kinesin-related pr
215	10	0.8	1190	2	S47536	oxysterol-binding
216	10	0.8	1195	2	S38174	probable purine nu
217	10	0.8	1228	2	S37621	Ca2+-transporting
218	10	0.8	1254	2	T18277	kinesin heavy chai
219	10	0.8	1294	2	T18473	hypothetical prote
220	10	0.8	1308	2	E71622	probable membrane
221	10	0.8	1365	2	T18419	hypothetical prote
222	10	0.8	1419	1	DVZQF	multidrug resistan
223	10	0.8	1438	2	B71610	WD40 WEB-1 homolog
224	10	0.8	1553	2	S67483	adenosinetriphosph
225	10	0.8	1576	2	T29237	hypothetical prote
226	10	0.8	1584	1	JN0114	kinesin-related pr
227	10	0.8	1584	2	T15822	kinesin-like prote
228	10	0.8	1711	2	T18429	hypothetical prote
229	10	0.8	1737	2	A59235	unconventional myo
230	10	0.8	1817	2	H71611	probable secreted
231	10	0.8	1893	2	T22661	hypothetical prote
232	10	0.8	1921	2	T13827	kinesin-73 - fruit
233	10	0.8	2359	2	T03094	A-kinase anchor pr
234	10	0.8	2529	2	A56923	transcription fact
235	10	0.8	2578	2	A56922	transcription fact
236	9	0.7	32	2	S08482	regulatory protein
237	9	0.7	114	2	T17927	hypothetical prote
238	9	0.7	131	2	S33375	homeotic protein A
239	9	0.7	160	2	S47661	topoisomerase II -
240	9	0.7	172	2	T20210	hypothetical prote
241	9	0.7	173	2	A47303	FTZ-F1 steroid rec
242	9	0.7	184	2	T28333	ORF MSV172 hypothe
243	9	0.7	257	2	T00406	hypothetical prote
244	9	0.7	257	2	E71601	probable integral
245	9	0.7	266	2	S61522	LMO protein - frui
246	9	0.7	300	2	H96636	hypothetical prote
247	9	0.7	305	2	T27151	hypothetical prote
248	9	0.7	321	2	G84750	probable homeodoma

249	9	0.7	323	2	T15311	hypothetical prote	322	9	0.7	1557	2	T13160	protein CNK - frui
250	9	0.7	344	2	T48827	hypothetical prote	323	9	0.7	1677	2	T43021	vitellogenin precu
251	9	0.7	372	2	T50014	trehalose-6-phosph	324	9	0.7	1714	2	E71609	Ser/Thr protein ki
252	9	0.7	372	2	T29359	hypothetical prote	325	9	0.7	2206	2	G71611	hypothetical prote
253	9	0.7	375	2	A46390	cAMP receptor subt	326	9	0.7	2485	1	H71621	serine/threonine-s
254	9	0.7	377	1	A48711	protein-tyrosine-p	327	9	0.7	2510	2	T28160	hypothetical prote
255	9	0.7	387	2	T25452	hypothetical prote	328	9	0.7	2657	2	T18497	hypothetical prote
256	9	0.7	398	2	D85080	KNAT1 homeobox-lik	329	9	0.7	2819	2	T09080	probable chloroqui
257	9	0.7	410	2	C86835	hypothetical prote	330	9	0.7	3092	2	S46009	GTPase-activating
258	9	0.7	452	2	T33049	hypothetical prote	331	9	0.7	3119	2	T18414	protein g377 - mal
259	9	0.7	472	2	S28026	nuclear pore compl	332	8	0.6	58	2	S78672	hypothetical prote
260	9	0.7	508	2	E71620	hypothetical prote	333	8	0.6	65	2	T17665	hypothetical prote
261	9	0.7	508	2	A32225	nerve growth facto	334	8	0.6	115	2	D61615	sericin MG-1 - gre
262	9	0.7	518	2	S23692	erythrocyte membra	335	8	0.6	117	2	A45509	desiccation-relate
263	9	0.7	528	2	S32593	[RNA-polymerase]-s	336	8	0.6	118	2	C41298	kinesin-like prote
264	9	0.7	564	2	S16304	benomyl/methotrexa	337	8	0.6	121	2	G71711	ribosomal protein
265	9	0.7	567	2	A71619	membrane transport	338	8	0.6	121	2	S67700	probable membrane
266	9	0.7	577	2	B44307	phosphoprotein pho	339	8	0.6	134	2	E84903	hypothetical prote
267	9	0.7	578	2	S55102	PSP2 protein - yea	340	8	0.6	145	2	H97707	30S ribosomal prot
268	9	0.7	581	2	F84599	probable kinesin h	341	8	0.6	147	2	JQ1320	high mobility grou
269	9	0.7	589	2	C85360	hypothetical prote	342	8	0.6	150	2	G64077	regulatory protein
270	9	0.7	594	2	JC5146	arylphorin gene-sp	343	8	0.6	151	2	T00286	hypothetical prote
271	9	0.7	598	2	T24717	hypothetical prote	344	8	0.6	163	2	T04467	hypothetical prote
272	9	0.7	602	2	B81420	GTP-binding protei	345	8	0.6	182	2	T20185	hypothetical prote
273	9	0.7	640	2	A41726	homeotic protein B	346	8	0.6	186	2	T30918	hypothetical prote
274	9	0.7	640	2	T03754	hypothetical prote	347	8	0.6	188	2	G96652	protein F23N19.15
275	9	0.7	642	2	S27806	homeotic protein B	348	8	0.6	205	2	T15450	hypothetical prote
276	9	0.7	643	2	S17997	hypothetical gene	349	8	0.6	215	2	F84508	probable CCAAT-box
277	9	0.7	658	2	C71856	excinuclease ABC c	350	8	0.6	236	2	B84718	hypothetical prote
278	9	0.7	658	2	B64659	excinuclease ABC c	351	8	0.6	237	2	T30920	homeobox protein -
279	9	0.7	669	2	S14535	asparagine-rich pr	352	8	0.6	242	2	T22915	hypothetical prote
280	9	0.7	675	2	B56262	excinuclease ABC c	353	8	0.6	256	2	S23691	erythrocyte membra
281	9	0.7	675	2	D81846	excinuclease ABC s	354	8	0.6	257	2	E89124	protein K07C11.1 [
282	9	0.7	675	2	F81095	excinuclease ABC c	355	8	0.6	259	2	G84685	probable DOF zinc
283	9	0.7	682	2	A49121	cell-surface molec	356	8	0.6	260	2	S60480	low temperature-in
284	9	0.7	682	2	A43318	connectin precurso	357	8	0.6	265	2	B86338	protein F5M15.22 [
285	9	0.7	704	2	S21911	BRcore-NS-Z3 prote	358	8	0.6	273	2	T03254	probable carbonate
286	9	0.7	708	2	T03835	vacA protein - sli	359	8	0.6	274	1	JQ0957	myb-related protei
287	9	0.7	713	1	S46838	hypothetical prote	360	8	0.6	275	2	D96826	hypothetical prote
288	9	0.7	719	2	S61046	ARP1 protein - yea	361	8	0.6	282	2	T08445	hypothetical prote
289	9	0.7	749	2	A57276	MAD1 protein - yea	362	8	0.6	284	2	S58650	hypothetical prote
290	9	0.7	770	1	A44337	kinesin-related pr	363	8	0.6	284	2	S27843	homeotic protein s
291	9	0.7	772	2	T43034	kinesin-like prote	364	8	0.6	285	2	C86423	unknown protein, 1
292	9	0.7	775	2	T43033	kinesin-like prote	365	8	0.6	286	2	C61615	sericin MG-2 - gre
293	9	0.7	779	1	S40382	box A-binding fact	366	8	0.6	294	2	S19226	cold-regulated pro
294	9	0.7	780	2	A48143	HF-1 regulatory el	367	8	0.6	298	2	A32872	myogenic factor CM
295	9	0.7	784	2	S26638	SPR-1 protein - hu	368	8	0.6	307	2	A45600	asparagine-rich bl
296	9	0.7	807	2	B71605	hypothetical prote	369	8	0.6	328	2	T00747	RING-H2 finger pro
297	9	0.7	832	2	T49494	condensin complex	370	8	0.6	329	1	S61884	carbonate dehydrat
298	9	0.7	834	2	T06055	hypothetical prote	371	8	0.6	330	1	S61883	carbonate dehydrat
299	9	0.7	860	2	JC5702	ErbB kinase activa	372	8	0.6	330	1	S61882	carbonate dehydrat
300	9	0.7	868	2	JC5701	ErbB kinase activa	373	8	0.6	331	1	S48675	carbonate dehydrat
301	9	0.7	883	2	T18484	hypothetical prote	374	8	0.6	333	2	T51650	probable transcrip
302	9	0.7	885	2	D86151	F22M8.8 protein -	375	8	0.6	334	2	B88482	protein C05D11.6 [
303	9	0.7	905	2	S56295	SAP155 protein - y	376	8	0.6	335	2	D86410	protein F3M18.16 [
304	9	0.7	928	2	S50578	hypothetical prote	377	8	0.6	336	2	S04682	ribosomal protein
305	9	0.7	960	1	A39651	discs-large tumor	378	8	0.6	339	2	F47301	VirB11 homolog - B
306	9	0.7	968	2	T45746	hypothetical prote	379	8	0.6	341	2	T22551	hypothetical prote
307	9	0.7	1003	2	T13856	ksr protein - frui	380	8	0.6	348	2	S19873	single-stranded DN
308	9	0.7	1070	2	T06733	kinesin homolog F2	381	8	0.6	354	2	A39609	glutathione transf
309	9	0.7	1091	2	S57112	JSN1 protein - yea	382	8	0.6	356	2	G84904	probable DOF zinc
310	9	0.7	1094	2	S46021	probable regulator	383	8	0.6	360	2	S69063	probable membrane
311	9	0.7	1110	2	T19673	hypothetical prote	384	8	0.6	360	2	T33835	hypothetical prote
312	9	0.7	1180	2	S69205	stripe a/b protein	385	8	0.6	361	2	C84613	probable cyclin D
313	9	0.7	1193	2	G71605	hypothetical prote	386	8	0.6	367	1	OZZQMY	circumsporozoite p
314	9	0.7	1221	2	T13283	probable transcrip	387	8	0.6	369	2	S76435	hypothetical prote
315	9	0.7	1254	2	T18425	hypothetical prote	388	8	0.6	372	2	I38042	single-stranded DN
316	9	0.7	1327	2	T14594	guanidine nucleoti	389	8	0.6	373	2	S72368	single-stranded DN
317	9	0.7	1397	2	T10466	DNA topoisomerase	390	8	0.6	376	2	T51673	myb-related transc
318	9	0.7	1407	2	B42239	adenylate cyclase	391	8	0.6	377	2	F84473	hypothetical prote
319	9	0.7	1436	2	D71618	hypothetical prote	392	8	0.6	383	2	S51651	cyclin delta-2 - A
320	9	0.7	1445	2	T14913	CAT8 protein - yea	393	8	0.6	384	2	AD3328	hflK protein [impo
321	9	0.7	1553	2	T18502	hypothetical prote	394	8	0.6	389	2	S53612	gene MSSP-2 protei

395	8	0.6	390	2	E81408	probable periplasm
396	8	0.6	397	2	T06531	pyruvate dehydroge
397	8	0.6	397	2	S54018	DBF2 protein-inter
398	8	0.6	403	2	T45580	hypothetical prote
399	8	0.6	403	2	T49480	hypothetical prote
400	8	0.6	403	2	S47659	CDC2/CDC13 suppres
401	8	0.6	411	2	E96665	protein F22C12.16
402	8	0.6	413	2	A55238	transcription fact
403	8	0.6	416	2	A55237	transcription fact
404	8	0.6	429	2	S24705	probable segment i
405	8	0.6	431	2	S50633	hypothetical prote
406	8	0.6	435	2	S51042	tat-binding protei
407	8	0.6	443	2	B84449	hypothetical prote
408	8	0.6	446	2	T14542	choline monooxygen
409	8	0.6	453	2	T47466	zinc finger protei
410	8	0.6	455	2	S46033	probable membrane
411	8	0.6	461	2	JN0097	secreted 45K prote
412	8	0.6	464	2	T45612	N-hydroxycinnamoyl
413	8	0.6	471	2	T32788	hypothetical prote
414	8	0.6	476	2	T17932	major capsid prote
415	8	0.6	479	2	T02623	hypothetical prote
416	8	0.6	485	2	S36184	translation elonga
417	8	0.6	486	2	F69762	transporter homolo
418	8	0.6	488	2	B90355	allantoin permease
419	8	0.6	488	2	B69825	cell wall-binding
420	8	0.6	492	2	E90371	allantoin permease
421	8	0.6	497	2	T06727	hypothetical prote
422	8	0.6	507	2	S05542	hypothetical prote
423	8	0.6	520	2	F86401	protein T22C5.11 [
424	8	0.6	532	2	T18432	hypothetical prote
425	8	0.6	535	2	T47629	phosphate transpor
426	8	0.6	536	2	H89130	protein F52E1.13b
427	8	0.6	538	2	T16034	hypothetical prote
428	8	0.6	554	2	S37936	regulatory protein
429	8	0.6	559	2	G82326	MSHA biogenesis pr
430	8	0.6	566	2	S54629	hypothetical prote
431	8	0.6	566	2	T49988	ovule development
432	8	0.6	567	2	S58750	NADH2 dehydrogenas
433	8	0.6	576	2	G96763	probable MAP kinas
434	8	0.6	581	2	H95214	ABC transporter, A
435	8	0.6	583	1	S20495	ferredoxin-nitrite
436	8	0.6	594	2	G98078	hypothetical prote
437	8	0.6	604	2	S58643	probable transcrip
438	8	0.6	605	2	T07123	nine-cis-epoxycaro
439	8	0.6	610	2	T18441	asparagine synthas
440	8	0.6	610	2	T19946	hypothetical prote
441	8	0.6	612	2	T08602	protein TipD - sli
442	8	0.6	628	2	T39494	kinesin-like prote
443	8	0.6	646	1	S16654	RNA-directed DNA p
444	8	0.6	653	2	B81277	hypothetical prote
445	8	0.6	655	2	A29945	neurogenesis regul
446	8	0.6	658	2	T41309	hypothetical threo
447	8	0.6	669	2	D82739	excinuclease ABC s
448	8	0.6	678	1	A30350	dorsal protein - f
449	8	0.6	688	2	T09941	transcription fact
450	8	0.6	695	2	T39954	probable gtpase ac
451	8	0.6	721	2	C84732	probable homeodoma
452	8	0.6	750	2	T04980	hypothetical prote
453	8	0.6	759	2	E84538	hypothetical prote
454	8	0.6	763	2	S17998	gene COX1 intron 4
455	8	0.6	769	2	T08431	gene bobby sox pro
456	8	0.6	779	2	T49717	related to BCS1 pr
457	8	0.6	784	1	A55236	kinesin-related pr
458	8	0.6	786	1	A47547	serine proteinease
459	8	0.6	840	2	S77615	hypothetical prote
460	8	0.6	841	2	S69563	suppressor protein
461	8	0.6	847	2	T12979	hypothetical prote
462	8	0.6	855	2	S47533	glucose-6-phosphat
463	8	0.6	859	2	S64195	HTR1 protein - yea
464	8	0.6	865	2	A25762	regulatory protein
465	8	0.6	870	2	B81689	ATP-dependent Clp
466	8	0.6	874	2	T00491	capsid protein pre
467	8	0.6	880	2	T38083	potassium transpor

468	8	0.6	883	2	T40128	kinesin-like prote
469	8	0.6	910	2	S40259	glucose-6-phosphat
470	8	0.6	911	2	T18451	hypothetical prote
471	8	0.6	941	2	T51135	ligand-gated chann
472	8	0.6	947	2	T03795	ornithine decarbox
473	8	0.6	962	2	D86186	hypothetical prote
474	8	0.6	962	2	S03818	carboxymethylcellu
475	8	0.6	970	2	F64230	spore germination
476	8	0.6	970	2	S48485	transcription acti
477	8	0.6	976	2	T51137	ionotropic glutama
478	8	0.6	1010	2	T13167	Lola-like protein
479	8	0.6	1017	2	T18488	hypothetical prote
480	8	0.6	1028	2	A56038	DNA-binding protei
481	8	0.6	1038	2	S38801	pullulanase (EC 3.
482	8	0.6	1043	2	T13733	FTZ-F1 protein - f
483	8	0.6	1060	2	T30347	vitellogenin conve
484	8	0.6	1067	2	S35423	protein kinase sgg
485	8	0.6	1099	2	T18257	phospholipase C -
486	8	0.6	1119	2	T18491	hypothetical prote
487	8	0.6	1123	2	T28139	PK4 protein kinase
488	8	0.6	1152	2	F86363	hypothetical prote
489	8	0.6	1163	2	D64315	type I restriction
490	8	0.6	1166	2	C96945	cation efflux syst
491	8	0.6	1169	2	AD1069	type 1 site-specif
492	8	0.6	1213	2	S16356	ovo protein - frui
493	8	0.6	1240	2	JC5209	insulin receptor s
494	8	0.6	1242	2	JS0670	insulin receptor s
495	8	0.6	1247	2	E71616	hypothetical prote
496	8	0.6	1278	2	A71609	probable secreted
497	8	0.6	1284	2	T13168	probable potassium
498	8	0.6	1314	1	TNBYR6	transcription regu
499	8	0.6	1367	2	H82874	conserved hypothet
500	8	0.6	1367	2	T18466	hypothetical prote
501	8	0.6	1384	2	S78132	DNA-directed RNA p
502	8	0.6	1398	2	H71606	hypothetical prote
503	8	0.6	1403	2	S24548	homeotic protein p
504	8	0.6	1422	2	T18404	chromatin remodell
505	8	0.6	1436	2	S57238	forked protein 5.4
506	8	0.6	1441	2	T18422	hypothetical prote
507	8	0.6	1449	2	S57237	forked protein 5.6
508	8	0.6	1468	2	A44345	nucleoporin - rat
509	8	0.6	1475	2	S42718	nuclear pore compl
510	8	0.6	1529	2	S69688	hypothetical prote
511	8	0.6	1597	1	BVFFSL	sol protein, large
512	8	0.6	1597	2	T08428	gene small optic 1
513	8	0.6	1704	2	T43141	vitellogenin 1 - m
514	8	0.6	1748	2	S42136	cnjB protein - Tet
515	8	0.6	1788	2	T31095	vitellogenin precu
516	8	0.6	1852	1	VJCH2	vitellogenin II pr
517	8	0.6	1965	2	T33216	hypothetical prote
518	8	0.6	2023	2	T13154	polycomb protein e
519	8	0.6	2098	2	T18397	protein CTRP - mal
520	8	0.6	2175	1	S03170	homeotic protein c
521	8	0.6	2391	2	T18410	carbamoyl-phosphat
522	8	0.6	2437	2	T18482	hypothetical prote
523	8	0.6	2500	2	G71609	hypothetical prote
524	8	0.6	2562	2	T14266	Xin protein - chic
525	8	0.6	3005	2	S33642	homeotic protein z
526	8	0.6	3190	2	T13828	CREB-binding prote
527	8	0.6	3759	2	A35085	trithorax protein
528	7	0.5	43	2	AG2078	hypothetical prote
529	7	0.5	46	2	B29184	vitellogenin - duc
530	7	0.5	47	2	H81576	hypothetical prote
531	7	0.5	54	2	T10786	conserved hypothet
532	7	0.5	54	2	T06514	hypothetical prote
533	7	0.5	68	2	F69799	hypothetical prote
534	7	0.5	70	2	C90886	hypothetical prote
535	7	0.5	70	2	D85732	hypothetical prote
536	7	0.5	70	2	B64898	hypothetical prote
537	7	0.5	76	2	S37191	hypothetical prote
538	7	0.5	77	2	T41800	AcMNPV orf55 - Bom
539	7	0.5	80	2	T27603	hypothetical prote
540	7	0.5	85	2	A86282	protein F10B6.20 [

541	7	0.5	89	2	B96930	transposon related
542	7	0.5	97	2	S06983	hypothetical prote
543	7	0.5	98	2	S13623	ht-en protein - le
544	7	0.5	99	2	T06083	probable auxin-ind
545	7	0.5	99	2	S69516	hypothetical prote
546	7	0.5	99	2	T49636	related to cell di
547	7	0.5	100	2	A26989	nonhistone chromos
548	7	0.5	101	2	S67034	probable membrane
549	7	0.5	102	2	T05499	auxin-induced prot
550	7	0.5	102	2	S72682	hypothetical prote
551	7	0.5	102	2	G96775	hypothetical prote
552	7	0.5	102	2	S26409	protein 108 precur
553	7	0.5	104	2	S55169	hypothetical prote
554	7	0.5	105	2	C89924	conserved hypothet
555	7	0.5	105	2	C86242	protein T16B5.1 [i
556	7	0.5	106	2	T14819	dehydrin DHN2 - so
557	7	0.5	107	2	B85356	glycine-rich prote
558	7	0.5	109	2	T12208	probable cytochrom
559	7	0.5	112	2	G72502	hypothetical prote
560	7	0.5	113	2	T48112	hypothetical prote
561	7	0.5	115	2	T07576	hypothetical prote
562	7	0.5	117	2	S48324	hypothetical prote
563	7	0.5	120	2	S65067	dehydrin DHN1 - po
564	7	0.5	121	2	E81872	hypothetical prote
565	7	0.5	123	2	S69460	hypothetical prote
566	7	0.5	124	2	S45788	probable membrane
567	7	0.5	125	2	S69871	hypothetical prote
568	7	0.5	127	1	KNMUHY	dehydrin-like prot
569	7	0.5	127	2	T45729	dehydrin-like prot
570	7	0.5	129	2	D41298	kinesin-like prote
571	7	0.5	129	2	T00282	hypothetical prote
572	7	0.5	132	2	T49536	hypothetical prote
573	7	0.5	133	2	A41298	kinesin-like prote
574	7	0.5	134	2	S51087	dhn1 protein - pot
575	7	0.5	135	2	T09876	dehydrin - upland
576	7	0.5	135	2	A42098	transcription fact
577	7	0.5	139	2	S05544	dehydrin 8 - barle
578	7	0.5	139	2	T05715	dehydrin - barley
579	7	0.5	140	2	T27059	hypothetical prote
580	7	0.5	142	2	T34136	hypothetical prote
581	7	0.5	143	2	S05543	dehydrin 9 - barle
582	7	0.5	145	2	S04042	embryonic abundant
583	7	0.5	149	2	S19130	rab15 protein - wh
584	7	0.5	149	2	T23179	hypothetical prote
585	7	0.5	150	2	H96756	hypothetical prote
586	7	0.5	150	2	S55882	CCHH finger protei
587	7	0.5	151	2	JQ2282	negatively phytoch
588	7	0.5	151	2	S11848	rab16D protein - r
589	7	0.5	153	2	S67294	hypothetical prote
590	7	0.5	154	2	T14727	dehydrin DHN1 - so
591	7	0.5	154	2	A05187	hypothetical prote
592	7	0.5	154	2	S55017	hypothetical prote
593	7	0.5	155	2	S43775	desiccation-relate
594	7	0.5	156	2	F85061	hypothetical prote
595	7	0.5	157	2	T10489	dehydrin 1 - Comme
596	7	0.5	158	2	A75556	hypothetical prote
597	7	0.5	159	2	T05279	transcription fact
598	7	0.5	159	2	T48956	hypothetical prote
599	7	0.5	160	2	B86614	CT832 hypothetical
600	7	0.5	160	2	T00836	probable dnaJ prot
601	7	0.5	160	2	F72010	CT832 hypothetical
602	7	0.5	161	2	S05547	dehydrin 17 - barl
603	7	0.5	161	2	T45055	hypothetical prote
604	7	0.5	161	2	D86418	unknown protein, 6
605	7	0.5	162	2	C85356	glycine-rich prote
606	7	0.5	163	2	S01152	RAB21 protein - ri
607	7	0.5	163	2	B81530	N utilization subs
608	7	0.5	164	2	T15525	hypothetical prote
609	7	0.5	164	2	S11846	rab16B protein - r
610	7	0.5	164	2	S11847	rab16C protein - r
611	7	0.5	164	2	T14187	hypothetical prote
612	7	0.5	164	2	T16168	hypothetical prote
613	7	0.5	165	2	T03392	probable dehydrin

614	7	0.5	166	2	T07089	dehydrin - soybean
615	7	0.5	166	2	F81707	hypothetical prote
616	7	0.5	166	2	S30833	hypothetical prote
617	7	0.5	167	2	S05545	dehydrin 3 - maize
618	7	0.5	167	2	T49638	hypothetical prote
619	7	0.5	167	2	AE2630	hypothetical prote
620	7	0.5	167	2	T47385	hypothetical prote
621	7	0.5	168	2	A39316	RAB-17 protein - m
622	7	0.5	168	2	S08633	RAB-17 protein - m
623	7	0.5	168	2	C75425	hypothetical prote
624	7	0.5	169	2	F84971	hypothetical prote
625	7	0.5	169	2	T34520	hypothetical prote
626	7	0.5	170	2	AF2621	polypeptide deform
627	7	0.5	170	2	F97403	polypeptide deform
628	7	0.5	171	2	T29161	hypothetical prote
629	7	0.5	172	2	G86463	hypothetical prote
630	7	0.5	172	2	T36107	probable serine/ar
631	7	0.5	173	2	F70112	hypothetical prote
632	7	0.5	174	2	F85028	hypothetical prote
633	7	0.5	177	2	A84856	hypothetical prote
634	7	0.5	177	2	F96739	hypothetical prote
635	7	0.5	177	2	T37212	hypothetical prote
636	7	0.5	178	2	T27193	hypothetical prote
637	7	0.5	178	2	T20599	hypothetical prote
638	7	0.5	179	2	B84587	probable glutaredo
639	7	0.5	180	2	A70170	hypothetical prote
640	7	0.5	180	2	C72676	hypothetical prote
641	7	0.5	181	2	G96715	transcription fact
642	7	0.5	181	2	T46017	hypothetical prote
643	7	0.5	182	2	G86410	protein F3M18.10 [
644	7	0.5	183	2	A56579	pre-intermoult gen
645	7	0.5	183	2	T06643	hypothetical prote
646	7	0.5	184	2	S12095	embryonic abundant
647	7	0.5	185	2	H84601	probable dehydrin
648	7	0.5	185	2	F96789	protein T23E18.12
649	7	0.5	185	2	T05008	hypothetical prote
650	7	0.5	187	2	S69466	hypothetical prote
651	7	0.5	187	2	T49491	hypothetical prote
652	7	0.5	188	1	SQRTAD	androgen-dependent
653	7	0.5	189	2	T02570	hypothetical prote
654	7	0.5	191	2	S29008	ADP-ribosylation f
655	7	0.5	192	2	S59537	heat shock transcr
656	7	0.5	192	2	H71605	probable integral
657	7	0.5	193	2	T19388	hypothetical prote
658	7	0.5	193	2	T00585	hypothetical prote
659	7	0.5	195	2	H96532	hypothetical prote
660	7	0.5	196	2	H82910	hypothetical prote
661	7	0.5	196	2	G84513	hypothetical prote
662	7	0.5	196	2	T46326	hypothetical prote
663	7	0.5	197	2	S55621	hypothetical prote
664	7	0.5	197	2	T26829	hypothetical prote
665	7	0.5	197	2	C88809	protein Y43C5B.1 [
666	7	0.5	197	2	A90365	hypothetical prote
667	7	0.5	197	2	A97060	N-terminal truncat
668	7	0.5	197	2	H85065	hypothetical prote
669	7	0.5	198	2	S07130	casein B precursor
670	7	0.5	198	2	A49243	GLORF-C4 protein -
671	7	0.5	199	2	H84015	hypothetical prote
672	7	0.5	199	2	T14468	pollen coat protei
673	7	0.5	200	2	JC4357	HMG1 protein - sea
674	7	0.5	200	2	S64905	hypothetical prote
675	7	0.5	200	2	T44143	DR3 protein [impor
676	7	0.5	200	2	T43953	hypothetical prote
677	7	0.5	200	2	T09783	dehydration-induci
678	7	0.5	202	2	D85058	hypothetical prote
679	7	0.5	202	2	T11744	dehydrin - kidney
680	7	0.5	203	2	T39117	hypothetical prote
681	7	0.5	203	2	S60449	probable membrane
682	7	0.5	204	2	B71433	hypothetical prote
683	7	0.5	205	2	T05713	dehydrin - barley
684	7	0.5	206	2	G85065	hypothetical prote
685	7	0.5	206	2	T23852	hypothetical prote
686	7	0.5	206	2	T34347	hypothetical prote

687	7	0.5	207	2	S58744	NADH2 dehydrogenas
688	7	0.5	207	2	S55200	hypothetical prote
689	7	0.5	207	2	E90140	hypothetical prote
690	7	0.5	209	2	E96517	hypothetical prote
691	7	0.5	209	2	T07779	dehydrin homolog C
692	7	0.5	210	2	B96694	hypothetical prote
693	7	0.5	210	2	T49785	hypothetical prote
694	7	0.5	211	2	D96507	hypothetical prote
695	7	0.5	211	2	T18799	hypothetical prote
696	7	0.5	211	2	E85091	hypothetical prote
697	7	0.5	212	2	T48580	hypothetical prote
698	7	0.5	213	2	S35522	heterochromatin pr
699	7	0.5	214	2	F84853	hypothetical prote
700	7	0.5	216	2	T29039	hypothetical prote
701	7	0.5	216	2	B96794	unknown protein F1
702	7	0.5	216	2	S50766	dehydrin-related p
703	7	0.5	216	2	S05575	sporozoite antigen
704	7	0.5	217	2	T06455	Myb26 protein - ga
705	7	0.5	218	2	E86750	hypothetical prote
706	7	0.5	218	2	JC7220	nuclear protein SR
707	7	0.5	220	2	S12849	coat protein TP4 -
708	7	0.5	221	2	T04271	probable thioredox
709	7	0.5	222	2	G86168	hypothetical prote
710	7	0.5	224	2	S78129	ribosomal protein
711	7	0.5	224	2	H72346	basal-body rod mod
712	7	0.5	225	2	S05546	dehydrin 18 - barl
713	7	0.5	225	2	B84653	TINY-like AP2 doma
714	7	0.5	225	2	S59852	DNA-binding protei
715	7	0.5	228	2	S19132	rab25 protein - ri
716	7	0.5	229	2	JQ0928	coat protein - tom
717	7	0.5	229	2	F84685	hypothetical prote
718	7	0.5	229	2	T04764	hypothetical prote
719	7	0.5	229	2	T52364	hypothetical prote
720	7	0.5	229	2	JC7219	nuclear protein SR
721	7	0.5	230	2	D86352	protein T26F17.14
722	7	0.5	230	2	S71747	DAG protein precur
723	7	0.5	230	2	T39598	hypothetical serin
724	7	0.5	230	2	T26896	hypothetical prote
725	7	0.5	231	2	S17367	rab15B protein - w
726	7	0.5	231	2	AI0091	probable RNA polym
727	7	0.5	233	2	T00966	hypothetical prote
728	7	0.5	233	2	H83753	hypothetical prote
729	7	0.5	233	2	H97094	hypothetical prote
730	7	0.5	234	2	T49448	hypothetical prote
731	7	0.5	235	2	S48924	hypothetical prote
732	7	0.5	235	2	D84602	hypothetical prote
733	7	0.5	235	2	T30656	hypothetical prote
734	7	0.5	238	2	S77699	inner cell wall ma
735	7	0.5	238	2	S66358	DNA-binding protei
736	7	0.5	238	2	T52505	hypothetical prote
737	7	0.5	238	2	T32889	hypothetical prote
738	7	0.5	238	2	A97336	Na+ ABC transporte
739	7	0.5	240	2	D84645	hypothetical prote
740	7	0.5	242	2	D86343	protein T22I11.13
741	7	0.5	243	2	T51989	ethylene responsiv
742	7	0.5	243	2	D84792	probable protein t
743	7	0.5	243	2	D84629	hypothetical prote
744	7	0.5	244	2	A42529	B29R protein - vac
745	7	0.5	244	2	A96806	hypothetical prote
746	7	0.5	244	2	T51481	hypothetical prote
747	7	0.5	244	2	B86197	hypothetical prote
748	7	0.5	244	2	E84885	hypothetical prote
749	7	0.5	244	2	S70176	yenR protein - Yer
750	7	0.5	244	2	T04700	hypothetical prote
751	7	0.5	244	2	D96937	SAM-dependent meth
752	7	0.5	245	2	AB0300	quorum-sensing tra
753	7	0.5	245	2	E86407	hypothetical prote
754	7	0.5	246	2	S58497	early auxin-induci
755	7	0.5	247	2	JE0307	membrane protein -
756	7	0.5	247	2	T47883	secretory protein-
757	7	0.5	249	2	E82917	methionine aminope
758	7	0.5	250	2	C71611	probable secreted
759	7	0.5	250	2	T10663	hypothetical prote

760	7	0.5	250	2	S36769	ubiquitin-protein
761	7	0.5	251	1	B64456	bioC protein homol
762	7	0.5	252	2	G84450	hypothetical prote
763	7	0.5	252	2	H84423	hypothetical prote
764	7	0.5	252	2	AE3276	hypothetical cytos
765	7	0.5	254	2	S50834	emerin - human
766	7	0.5	255	1	WJHU4B	homeotic protein H
767	7	0.5	255	2	AB1942	hypothetical prote
768	7	0.5	255	2	T05958	cold acclimation p
769	7	0.5	257	2	G84712	hypothetical prote
770	7	0.5	258	1	WMVZ35	35K major secreted
771	7	0.5	258	2	S71561	drought-induced pr
772	7	0.5	258	2	T05194	hypothetical prote
773	7	0.5	259	2	A84476	hypothetical prote
774	7	0.5	259	2	S55884	CCHH finger protei
775	7	0.5	259	2	T47627	hypothetical prote
776	7	0.5	259	2	AB1080	conserved hypothet
777	7	0.5	259	2	S64423	probable membrane
778	7	0.5	260	2	B64563	hypothetical prote
779	7	0.5	260	2	D86266	hypothetical prote
780	7	0.5	260	2	B96686	probable C2H2-type
781	7	0.5	261	2	T06650	myb-related protei
782	7	0.5	261	2	E85439	thiol-disulfide in
783	7	0.5	261	2	T09075	hypothetical prote
784	7	0.5	262	2	T40941	hypothetical serin
785	7	0.5	262	2	T02203	finger protein Dof
786	7	0.5	262	2	T14807	hypothetical prote
787	7	0.5	262	2	S43953	cold acclimation p
788	7	0.5	263	2	H84918	glutathione S-tran
789	7	0.5	263	2	S05433	corticotropin / li
790	7	0.5	263	2	JC5271	oxygen-evolving co
791	7	0.5	263	2	H84547	hypothetical prote
792	7	0.5	264	2	T09377	hypothetical prote
793	7	0.5	264	2	T10637	hypothetical prote
794	7	0.5	265	2	S48360	hypothetical prote
795	7	0.5	265	2	D96667	hypothetical prote
796	7	0.5	265	2	F85014	hypothetical prote
797	7	0.5	265	2	T29973	hypothetical prote
798	7	0.5	268	2	T05123	hypothetical prote
799	7	0.5	268	2	T06802	cold acclimation p
800	7	0.5	269	2	C84707	hypothetical prote
801	7	0.5	270	2	F86177	protein F19P19.1 [
802	7	0.5	270	2	T22151	hypothetical prote
803	7	0.5	271	2	T13013	hypothetical prote
804	7	0.5	273	2	B64446	formylmethanofuran
805	7	0.5	273	2	F71645	hypothetical prote
806	7	0.5	274	2	S26953	photosystem II 22K
807	7	0.5	274	2	T07393	myb-related transc
808	7	0.5	275	2	T05595	hypothetical prote
809	7	0.5	276	2	A31864	vitellogenin III -
810	7	0.5	276	2	H97861	hypothetical prote
811	7	0.5	276	2	F95259	conserved hypothet
812	7	0.5	276	2	A99525	conserved hypothet
813	7	0.5	277	2	T52513	hypothetical prote
814	7	0.5	277	2	T19378	hypothetical prote
815	7	0.5	278	2	T51641	myb-related transc
816	7	0.5	279	2	T01924	hypothetical prote
817	7	0.5	279	2	A86312	F11A6.5 protein -
818	7	0.5	279	2	H89827	mevalonate kinase
819	7	0.5	280	1	EDBE3T	30K immediate-earl
820	7	0.5	281	2	T16813	hypothetical prote
821	7	0.5	281	2	T01612	hypothetical prote
822	7	0.5	282	2	C96687	hypothetical prote
823	7	0.5	282	2	S61663	probable membrane
824	7	0.5	282	2	T21222	hypothetical prote
825	7	0.5	283	2	F86450	hypothetical prote
826	7	0.5	284	1	KART	alpha-casein precu
827	7	0.5	284	2	A35419	neutrophil protein
828	7	0.5	285	2	T06627	hypothetical prote
829	7	0.5	285	2	G70100	hypothetical prote
830	7	0.5	285	2	T18689	hypothetical prote
831	7	0.5	286	2	H90018	hypothetical prote
832	7	0.5	286	2	C84857	hypothetical prote

833	7	0.5	286	2	T48572	hypothetical prote
834	7	0.5	286	2	T52373	homeobox protein T
835	7	0.5	288	2	B40722	homeotic protein 1
836	7	0.5	289	2	T03395	probable lipase -
837	7	0.5	289	2	T52354	hypothetical prote
838	7	0.5	290	2	T21868	hypothetical prote
839	7	0.5	291	2	T08189	hypothetical prote
840	7	0.5	292	2	I51171	transcription fact
841	7	0.5	292	2	T48327	hypothetical prote
842	7	0.5	292	2	S41012	hypothetical prote
843	7	0.5	294	2	T05381	hypothetical prote
844	7	0.5	294	2	S38983	kinesin-related pr
845	7	0.5	294	2	A96707	hypothetical prote
846	7	0.5	294	2	T08408	transcription fact
847	7	0.5	295	2	B86371	hypothetical prote
848	7	0.5	295	2	T00399	probable AP2 domai
849	7	0.5	295	2	T23989	hypothetical prote
850	7	0.5	297	2	E84731	hypothetical prote
851	7	0.5	298	2	A86840	hypothetical prote
852	7	0.5	299	2	F84810	hypothetical prote
853	7	0.5	300	2	C84853	hypothetical prote
854	7	0.5	302	2	T01942	hypothetical prote
855	7	0.5	302	2	T01540	hypothetical prote
856	7	0.5	302	2	C86480	33.2K hypothetical
857	7	0.5	302	2	H86391	hypothetical prote
858	7	0.5	303	2	T15694	hypothetical prote
859	7	0.5	303	2	B84744	hypothetical prote
860	7	0.5	303	2	T04681	hypothetical prote
861	7	0.5	304	2	T09357	hypothetical prote
862	7	0.5	304	2	T16535	hypothetical prote
863	7	0.5	305	2	T47844	hypothetical prote
864	7	0.5	305	2	JC4525	nucleic acid-bindi
865	7	0.5	306	2	S33538	embryogenic cell p
866	7	0.5	306	2	T47832	hypothetical prote
867	7	0.5	308	2	C84768	hypothetical prote
868	7	0.5	310	2	JC4343	uridine phosphoryl
869	7	0.5	310	2	D86423	hypothetical prote
870	7	0.5	311	2	T23846	hypothetical prote
871	7	0.5	313	2	S55899	transcription elon
872	7	0.5	313	2	S59448	hypothetical prote
873	7	0.5	313	2	S01303	homeotic protein a
874	7	0.5	313	2	T05256	hypothetical prote
875	7	0.5	314	2	T08933	hypothetical prote
876	7	0.5	314	2	A41544	developmental prot
877	7	0.5	314	2	A36195	developmental prot
878	7	0.5	314	2	T48514	hypothetical prote
879	7	0.5	315	2	JC7572	somite Maf1 protei
880	7	0.5	316	2	E96770	hypothetical prote
881	7	0.5	317	2	T39869	probable lysophosp
882	7	0.5	318	2	T32340	hypothetical prote
883	7	0.5	318	2	T20063	hypothetical prote
884	7	0.5	319	2	T43040	hypothetical prote
885	7	0.5	319	2	T25666	hypothetical prote
886	7	0.5	319	2	B86433	hypothetical prote
887	7	0.5	320	2	D86259	protein T12C24.6 [
888	7	0.5	321	2	A84792	hypothetical prote
889	7	0.5	321	2	S55640	hypothetical prote
890	7	0.5	323	2	I51751	homeotic protein o
891	7	0.5	323	2	G96544	hypothetical prote
892	7	0.5	323	2	F86192	hypothetical prote
893	7	0.5	324	2	B69521	hypothetical prote
894	7	0.5	324	2	C86426	hypothetical prote
895	7	0.5	325	2	G96718	unknown protein, 5
896	7	0.5	325	2	S46760	hypothetical prote
897	7	0.5	326	2	C96735	unknown protein F2
898	7	0.5	327	2	B84781	hypothetical prote
899	7	0.5	328	1	S10200	carbonate dehydrat
900	7	0.5	328	2	G02469	homeotic protein D
901	7	0.5	329	2	D64245	peripheral membran
902	7	0.5	329	2	T13016	hypothetical prote
903	7	0.5	332	2	C48835	kinesin-like prote
904	7	0.5	332	2	T48083	hypothetical prote
905	7	0.5	333	1	JW0097	bicoid-related hom
906	7	0.5	335	2	B84024	phosphate ABC tran
907	7	0.5	336	1	S28412	carbonate dehydrat
908	7	0.5	338	2	D72327	heat shock operon
909	7	0.5	339	2	T04900	hypothetical prote
910	7	0.5	339	2	D90106	hypothetical prote
911	7	0.5	339	2	H86210	hypothetical prote
912	7	0.5	339	2	T28178	hypothetical prote
913	7	0.5	342	2	AF3357	anthranilate phosp
914	7	0.5	343	2	G69654	cell wall lytic ac
915	7	0.5	343	2	T01374	hypothetical prote
916	7	0.5	343	2	T13995	gag protein - frui
917	7	0.5	344	2	T05064	hypothetical prote
918	7	0.5	344	2	T40167	hypothetical prote
919	7	0.5	345	2	T06098	hypothetical prote
920	7	0.5	345	2	T07777	probable homeobox
921	7	0.5	347	2	T07108	glucan endo-1,3-be
922	7	0.5	347	2	H96504	hypothetical prote
923	7	0.5	347	2	H84608	probable RNA-bindi
924	7	0.5	348	2	T28310	hypothetical prote
925	7	0.5	349	2	T41394	hypothetical serin
926	7	0.5	349	2	T29931	hypothetical prote
927	7	0.5	350	2	T07061	L-lactate dehydrog
928	7	0.5	350	2	T06689	heat shock protein
929	7	0.5	352	2	S22464	strictosidine synt
930	7	0.5	353	2	AB2396	hypothetical prote
931	7	0.5	354	2	G71403	hypothetical prote
932	7	0.5	354	2	T49806	hypothetical prote
933	7	0.5	355	2	T04317	homeobox protein L
934	7	0.5	355	2	T38483	hypothetical prote
935	7	0.5	355	2	T24010	hypothetical prote
936	7	0.5	357	2	A54512	serine-repeat anti
937	7	0.5	358	2	C86291	hypothetical prote
938	7	0.5	360	2	T45649	hypothetical prote
939	7	0.5	362	2	T51464	RING-H2 zinc finge
940	7	0.5	363	2	T49209	leucoanthocyanidin
941	7	0.5	363	2	S60684	adenylate cyclase
942	7	0.5	363	2	A96573	protein F12M16.15
943	7	0.5	364	2	A84890	probable flavonol
944	7	0.5	364	2	I48188	gene NKx6.1 protei
945	7	0.5	364	2	A96730	hypothetical prote
946	7	0.5	364	2	H86384	40.1K hypothetical
947	7	0.5	365	2	C72308	conserved hypothet
948	7	0.5	365	2	F85018	hypothetical prote
949	7	0.5	366	2	T10452	farnesyltranstrans
950	7	0.5	366	2	T00439	hypothetical prote
951	7	0.5	366	2	A86392	hypothetical prote
952	7	0.5	367	2	E64399	hypothetical prote
953	7	0.5	367	2	S28604	St12p protein - Ar
954	7	0.5	368	2	F69531	ATP-dependent RNA
955	7	0.5	368	2	A54430	hypoxic function t
956	7	0.5	368	2	T51200	hypothetical prote
957	7	0.5	369	2	I38659	melanoma antigen M
958	7	0.5	370	2	S59539	heat shock transcr
959	7	0.5	371	2	T06382	Knox protein 1 - g
960	7	0.5	372	2	E85170	hypothetical prote
961	7	0.5	372	2	S20056	para-hydroxybenzoa
962	7	0.5	375	2	T06436	probable ornithine
963	7	0.5	375	2	E96567	hypothetical prote
964	7	0.5	376	2	C85155	RING-H2 finger pro
965	7	0.5	376	2	S45763	hypothetical prote
966	7	0.5	377	1	W2WL13	E2 protein - human
967	7	0.5	377	1	MFNZP2	matrix protein - p
968	7	0.5	377	2	C64404	hypothetical prote
969	7	0.5	379	2	S42543	hypothetical prote
970	7	0.5	380	2	F89811	cystathionine gamm
971	7	0.5	381	2	T48623	hypothetical prote
972	7	0.5	381	2	H86856	hypothetical prote
973	7	0.5	381	2	C96657	hypothetical prote
974	7	0.5	381	2	T09640	protein phosphatas
975	7	0.5	382	2	F75179	DNA topoisomerase
976	7	0.5	382	2	D84766	hypothetical prote
977	7	0.5	383	2	T38443	hypothetical prote
978	7	0.5	383	2	T38442	hypothetical prote

979	7	0.5	383	2	T06753	zinc finger protei
980	7	0.5	383	2	T46707	proteophosphoglyca
981	7	0.5	383	2	S67462	hypothetical prote
982	7	0.5	383	2	S61195	hypothetical prote
983	7	0.5	383	2	A84552	probable thioredox
984	7	0.5	383	2	AE2295	hypothetical prote
985	7	0.5	384	2	G96947	cystathionine gamm
986	7	0.5	386	2	S58755	ribosomal protein
987	7	0.5	386	2	F90112	DEAD box protein [
988	7	0.5	386	2	T48158	hypothetical prote
989	7	0.5	386	2	A82284	conserved hypothet
990	7	0.5	387	2	C96720	hypothetical prote
991	7	0.5	389	2	T04816	myb-related protei
992	7	0.5	391	2	S18666	KIN17 protein - mo
993	7	0.5	392	2	S11998	finger protein odd
994	7	0.5	392	2	T43490	hypothetical prote
995	7	0.5	393	2	A45017	transcription fact
996	7	0.5	393	2	T00647	glycosyl transfera
997	7	0.5	393	2	B84425	probable protein t
998	7	0.5	393	2	T48907	St12p protein [imp
999	7	0.5	394	2	C84905	probable extensin
1000	7	0.5	395	2	T00516	hypothetical prote

ALIGNMENTS

```
RESULT 1
S14577
asparagine-rich protein (clone 14C1) - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jun-2000
C;Accession: S14577
R;Schreiber, L.; Deutsche, U.; Storck, T.; Mueller-Hill, D.
submitted to the EMBL Data Library, December 1989
A;Reference number: S14469
A;Accession: S14577
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-391 <SCH>
A;Cross-references: EMBL:X17483; NID:g9842; PID:g9843
C;Superfamily: Saccharomyces cerevisiae ribosomal protein var1

  Query Match             1.3%;  Score 17;  DB 2;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db        162 TNNNNNNNNNNNNNNNN 178

RESULT 2
S14508
asparagine-rich protein (clone 52C11) - Plasmodium falciparum (fragment)
C;Species: Plasmodium falciparum
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Sep-1997
C;Accession: S14508
R;Schreiber, L.; Deutsche, U.; Storck, T.; Mueller-Hill, D.
submitted to the EMBL Data Library, December 1989
A;Reference number: S14469
A;Accession: S14508
A;Molecule type: mRNA
A;Residues: 1-419 <SCH>
A;Cross-references: EMBL:X17489; NID:g9854; PID:g9855

  Query Match             1.3%;  Score 17;  DB 2;  Length 419;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db         25 TNNNNNNNNNNNNNNNN 41

RESULT 3
S50452
hypothetical protein YEL007w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 19-Apr-2002
C;Accession: S50452
R;Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A;Description: Saccharomyces cerevisiae chromosome V cosmids 9871, 8199, 9867, 9495 and
A;Reference number: S50428
A;Accession: S50452
A;Molecule type: DNA
A;Residues: 1-666 <DIE>
A;Cross-references: EMBL:U18530; NID:g602367; PID:g602374; GSPDB:GN00005; MIPS:YEL007w
C;Genetics:
A;Gene: SGD:TOS9; MIPS:YEL007w
A;Cross-references: SGD:S0000733
A;Map position: 5L

  Query Match             1.3%;  Score 17;  DB 2;  Length 666;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        351 NNNNNNNNNNNNNNNNI 367

RESULT 4
S23467
probable long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) - malaria parasite (Plasmodium f
C;Species: Plasmodium falciparum
C;Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 03-Jun-2002
C;Accession: S23467
R;Zhao, Y.; Kappes, B.; Yang, J.; Franklin, R.M.
Eur. J. Biochem. 207, 305-313, 1992
A;Title: Molecular cloning, stage-specific expression and cellular distribution of a pu
A;Reference number: S23466; MUID:92331669; PMID:1378403
A;Accession: S23467
A;Status: preliminary; translation not shown
A;Molecule type: mRNA
A;Residues: 1-749 <ZHA>
A;Cross-references: EMBL:X63648; NID:g9937; PID:g9939
C;Superfamily: human long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homology
C;Keywords: acid-thiol ligase; coenzyme A
F;103-732/Domain: acetate-CoA ligase homology <ACL>

  Query Match             1.3%;  Score 17;  DB 2;  Length 749;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        271 NNNNNNNNNNNNNNNNI 287

RESULT 5
S45825
cell division control protein CDC27 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YBL0718; protein YBL084c
C;Species: Saccharomyces cerevisiae
C;Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 11-Jan-2000
C;Accession: S45825; S45422; S28742; S59219
R;Domdey, H.; Gassenhuber, H.; Obermaier, B.; Piravandi, E.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45816
A;Accession: S45825
A;Molecule type: DNA
A;Residues: 1-758 <DOM>
A;Cross-references: EMBL:Z35845; NID:g536135; PIDN:CAA84905.1; PID:g536136; MIPS:YBL084
A;Experimental source: strain S288C
```

```
R;Obermaier, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.
submitted to the EMBL Data Library, May 1994
A;Description: Sequence analysis of a 78,6 kb segment of the left end of Saccaromyces ce
A;Reference number: S45387
A;Accession: S45422
A;Molecule type: DNA
A;Residues: 141-758 <OBE>
A;Cross-references: EMBL:X79489; NID:g496661; PIDN:CAA56022.1; PID:g496695
A;Experimental source: strain S288C
R;Sikorski, R.S.; Michaud, W.A.; Wootton, J.C.; Boguski, M.S.; Connelly, C.; Hieter, P.
Cold Spring Harb. Symp. Quant. Biol. 56, 663-673, 1991
A;Title: TPR proteins as essential components of the yeast cell cycle.
A;Reference number: S28742; MUID:92306611; PMID:1819514
A;Accession: S28742
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 607-619 <SIK>
R;Obermaier, B.; Gassenhuber, J.; Piravandi, E.; Domdey, H.
Yeast 11, 1103-1112, 1995
A;Title: Sequence analysis of a 78.6 kb segment of the left end of Saccharomyces cerevis
A;Reference number: S59184; MUID:96076635; PMID:7502586
A;Accession: S59219
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 442-691 <OBW>
A;Cross-references: EMBL:X79489
C;Genetics:
A;Gene: SGD:CDC27; SNB1
A;Cross-references: SGD:S0000180; MIPS:YBL084c
A;Map position: 2L
C;Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat ho
C;Keywords: nucleus
F;472-505/Domain: tetratricopeptide repeat homology <TT1>
F;506-539/Domain: tetratricopeptide repeat homology <TT2>
F;540-573/Domain: tetratricopeptide repeat homology <TT3>
F;574-607/Domain: tetratricopeptide repeat homology <TT4>
F;608-641/Domain: tetratricopeptide repeat homology <TT5>
F;642-675/Domain: tetratricopeptide repeat homology <TT6>
F;676-709/Domain: tetratricopeptide repeat homology <TT7>
F;710-743/Domain: tetratricopeptide repeat homology <TT8>

  Query Match             1.3%;  Score 17;  DB 2;  Length 758;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        373 NNNNNNNNNNNNNNNNI 389

RESULT 6
S68824
rngB protein, cytosolic - slime mold (Dictyostelium discoideum)
N;Alternate names: RING finger protein
C;Species: Dictyostelium discoideum
C;Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 24-Apr-1998
C;Accession: S68824
R;Kawata, T.; Steel, J.B.; Williams, J.G.
FEBS Lett. 386, 103-109, 1996
A;Title: RNGB: a Dictyostelium RING finger protein that is specifically located in matur
A;Reference number: S68824; MUID:96228044; PMID:8647262
A;Accession: S68824
A;Molecule type: mRNA
A;Residues: 1-943 <KAW>
A;Experimental source: maturing cells
C;Genetics:
A;Gene: rngB
C;Superfamily: RING finger homology
C;Keywords: zinc
F;893-937/Domain: RING finger homology <RNG>

  Query Match             1.3%;  Score 17;  DB 2;  Length 943;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db        433 TNNNNNNNNNNNNNNNN 449

RESULT 7
T18275
1-phosphatidylinositol 3-kinase (EC 2.7.1.137) 4 - slime mold (Dictyostelium discoideum
C;Species: Dictyostelium discoideum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18275
R;Zhou, K.; Takegawa, K.; Emr, S.D.; Firtel, R.A.
Mol. Cell. Biol. 15, 5645-5656, 1995
A;Title: A phosphatidylinositol (PI) kinase gene family in Dictyostelium discoideum: Bi
A;Reference number: Z06411
A;Accession: T18275
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1093 <ZHO>
A;Cross-references: EMBL:U23479; NID:g733527; PID:g733528; PIDN:AAA85725.1
C;Genetics:
A;Note: PIK4
C;Keywords: phosphotransferase

  Query Match             1.3%;  Score 17;  DB 2;  Length 1093;
  Best Local Similarity   100.0%;  Pred. No. 2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        282 NNNNNNNNNNNNNNNNI 298

RESULT 8
S71092
UV-damaged DNA binding protein repE - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 01-Dec-2000
C;Accession: S71092; S21443
R;Alexander, H.; Lee, S.K.; Yu, S.L.; Alexander, S.
Nucleic Acids Res. 24, 2295-2301, 1996
A;Title: repE-The Dictyostelium homolog of the human xeroderma pigmentosum group E gene
A;Reference number: S71092; MUID:96279729; PMID:8710499
A;Accession: S71092
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1139 <ALE>
A;Cross-references: EMBL:U50042; NID:g1399511; PIDN:AAC47162.1; PID:g1399512
R;Sydow, L.; Alexander, H.; Alexander, S.
submitted to the EMBL Data Library, April 1992
A;Reference number: S21443
A;Accession: S21443
A;Molecule type: mRNA
A;Residues: 860-1139 <SYD>
A;Cross-references: EMBL:X65937
C;Genetics:
A;Gene: repE
A;Introns: 654/2
C;Superfamily: UV-damaged DNA-binding protein

  Query Match             1.3%;  Score 17;  DB 2;  Length 1139;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        115 NNNNNNNNNNNNNNNNI 131

RESULT 9
S57698
```

```
regulatory protein PHO81 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein 1178; protein G8567; protein YGR233c
C;Species: Saccharomyces cerevisiae
C;Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 23-Mar-2001
C;Accession: S57698; S20136; S41074; JC1032; S64557; S63914
R;van der Aart, Q.J.M.; Kleine, K.; Steensma, H.Y.
submitted to the EMBL Data Library, June 1995
A;Description: Sequence analysis of the 43 KB CRM1-YLM9-PET54-SMI1-PHO81-YHB4-PFK1 regic
A;Reference number: S57680
A;Accession: S57698
A;Molecule type: DNA
A;Residues: 1-1178 <VAN>
A;Cross-references: EMBL:X87941; NID:g886908; PID:g886927
A;Experimental source: strain S288C
R;Coche, T.; Prozzi, D.; Legrain, M.; Hilger, F.; Vandenhaute, J.
Nucleic Acids Res. 18, 2176, 1990
A;Title: Nucleotide sequence of the PHO81 gene involved in the regulation of the repress
A;Reference number: S20136; MUID:90245670; PMID:2186378
A;Accession: S20136
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-918,920-1178 <COC>
A;Cross-references: EMBL:X52482; NID:g4139; PIDN:CAA36726.1; PID:g4140
A;Experimental source: strain GRF 88
R;Creasy, C.L.; Madden, S.L.; Bergman, L.W.
Nucleic Acids Res. 21, 1975-1982, 1993
A;Title: Molecular analysis of the PHO81 gene of Saccharomyces cerevisiae.
A;Reference number: S41074; MUID:93261834; PMID:8493108
A;Accession: S41074
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-10 <CRE>
A;Cross-references: GB:S61041; NID:g385415; PIDN:AAD13922.1; PID:g4261622
R;Wu, J.S.; Ao, S.Z.
Acta Biochim. Biophys. Sin. 26, 283-287, 1994
A;Title: Conformation prediction and function analysis of yeast PHO81 protein.
A;Reference number: JC1032
A;Accession: JC1032
A;Molecule type: DNA
A;Residues: 1-39,41-78,80-119,121-159,161-199,201-230,232-279,281-282,'L',284-319,321-35
79,881-892,'M',894-919,922-960,962-1000,1002-1040,1042-1080,1082-1103,'L',1105-1120,1122
R;van der Aart, Q.J.M.; Steensma, H.Y.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64541
A;Accession: S64557
A;Molecule type: DNA
A;Residues: 1-1178 <VAW>
A;Cross-references: EMBL:Z73018; NID:g1323420; PID:g1323421; MIPS:YGR233c
A;Experimental source: strain S288C
R;van der Aart, Q.J.M.; Kleine, K.; Steensma, H.Y.
Yeast 12, 385-390, 1996
A;Title: Sequence analysis of the 43 kb CRM1-YLM9-PET54-DIE2-SMI1-PHO81-YHB4-PFK1 region
A;Reference number: S63896; MUID:96267763; PMID:8701610
A;Accession: S63914
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1178 <VAF>
A;Cross-references: EMBL:X87941; NID:g886908; PIDN:CAA61183.1; PID:g886927
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995
C;Genetics:
A;Gene: SGD:PHO81
A;Cross-references: SGD:S0003465; MIPS:YGR233c
A;Map position: 7R
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: transmembrane protein
F;330-346/Domain: transmembrane #status predicted <TM1>
F;904-920/Domain: transmembrane #status predicted <TM2>

  Query Match             1.3%;  Score 17;  DB 2;  Length 1178;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        233 NNNNNNNNNNNNNNNNI 249

RESULT 10
T18287
protein-tyrosine kinase (EC 2.7.1.112) - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Feb-2000
C;Accession: T18287
R;Adler, K.
submitted to the EMBL Data Library, July 1996
A;Reference number: Z18856
A;Accession: T18287
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1338 <ADL>
A;Cross-references: EMBL:U64830; NID:g1468982; PID:g1468983; PIDN:AAB04999.1
C;Genetics:
A;Introns: 1181/3
C;Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase

  Query Match             1.3%;  Score 17;  DB 2;  Length 1338;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db          5 NNNNNNNNNNNNNNNNI 21

RESULT 11
T14577
protein kinase YakA (EC 2.7.1.-) - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14577
R;Kuspa, A.; Lu, S.; Souza, G.M.
submitted to the EMBL Data Library, January 1998
A;Description: YakA, a protein kinase required for the growth to development transition
A;Reference number: Z18146
A;Accession: T14577
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1457 <KUS>
A;Cross-references: EMBL:AF045453; NID:g2854116; PID:g2854117; PIDN:AAC02554.1
C;Genetics:
A;Gene: yakA
C;Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protein k

  Query Match             1.3%;  Score 17;  DB 2;  Length 1457;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db         44 TNNNNNNNNNNNNNNNN 60

RESULT 12
S71628
sensory transduction histidine kinase dokA - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 11-Jan-2002
C;Accession: S71628; S78068
R;Schuster, S.C.; Noegel, A.A.; Oehme, F.; Gerisch, G.; Simon, M.I.
EMBO J. 15, 3880-3889, 1996
A;Title: The hybrid histidine kinase DokA is part of the osmotic response system of Dic
A;Reference number: S71628; MUID:96324396; PMID:8670893
A;Accession: S71628
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1670 <SCH>
```

```
A;Cross-references: EMBL:X96869
A;Experimental source: strain AX2; substrain 214
R;Schuster, S.C.; Noegel, A.A.; Oehme, F.; Gerisch, G.; Simon, M.I.
submitted to the EMBL Data Library, March 1996
A;Description: The hybrid histidine kinase DokA is part of the osmotic response system c
A;Reference number: S78068
A;Accession: S78068
A;Molecule type: DNA
A;Residues: 1-149,'E',151-219,'TRVLKLIQSTNNWIYWY',238-1671 <SCW>
A;Cross-references: EMBL:X96869; NID:g1237201; PIDN:CAA65612.1; PID:g1237202
C;Genetics:
A;Gene: dokA
C;Function:
A;Description: modulates cell response to changes in osmolarity; involved in spore forma
C;Superfamily: response regulator homology
C;Keywords: phosphoprotein; signal transduction
F;1520-1629/Domain: response regulator homology <RRH2>
F;1568/Binding site: phosphate (Asp) (covalent) #status predicted

  Query Match             1.3%;  Score 17;  DB 2;  Length 1671;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNNN 757
              |||||||||||||||||
Db        374 TNNNNNNNNNNNNNNNN 390


RESULT 13
T08875
histidine kinase homolog DHKB - slime mold (Dictyostelium discoideum)
N;Alternate names: hybrid histidine kinase DHKB
C;Species: Dictyostelium discoideum
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-May-2000
C;Accession: T08875
R;Zinda, M.J.; Singleton, C.K.
Dev. Biol. 196, 171-183, 1998
A;Title: The hybrid histidine kinase dhkB regulates spore germination in Dictyostelium d
A;Reference number: Z16506; MUID:98248997; PMID:9576830
A;Accession: T08875
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1969 <SIN>
A;Cross-references: EMBL:AF024654; NID:g2460282; PID:g2460283
A;Experimental source: strain KAx3
C;Genetics:
A;Gene: dhkB
A;Introns: 790/3
C;Superfamily: response regulator homology
C;Keywords: protein kinase; transmembrane protein
F;1841-1964/Domain: response regulator homology <RRH>

  Query Match             1.3%;  Score 17;  DB 2;  Length 1969;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db         37 NNNNNNNNNNNNNNNNI 53


RESULT 14
T18440
hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18440
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18440
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
```

```
A;Residues: 1-4550 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB11121.1
C;Genetics:
A;Map position: 3
A;Note: C0425w

  Query Match             1.3%;  Score 17;  DB 2;  Length 4550;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db       1517 NNNNNNNNNNNNNNNNI 1533


RESULT 15
T02389
hypothetical protein At2g44340 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F4I1.15
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C;Accession: T02389; C84877
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kau
submitted to the EMBL Data Library, May 1998
A;Description: Arabidopsis thaliana chromosome II BAC F4I1 genomic sequence.
A;Reference number: Z14667
A;Accession: T02389
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-188 <ROU>
A;Cross-references: EMBL:AC004521; NID:g3128166; PID:g3128179
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84877
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-188 <STO>
A;Cross-references: GB:AE002093; NID:g3128179; PIDN:AAC16083.1; GSPDB:GN00139
C;Genetics:
A;Gene: F4I1.15; At2g44340
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T10K17.210

  Query Match             1.2%;  Score 16;  DB 2;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 4e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        161 NNNNNNNNNNNNNNNN 176


RESULT 16
S05356
hypothetical protein (clone pLK330) - slime mold (Dictyostelium discoideum) (fragment)
C;Species: Dictyostelium discoideum
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 29-Oct-1999
C;Accession: S05356
R;Shaw, D.R.; Richter, H.; Giorda, R.; Ohmachi, T.; Ennis, H.L.
Mol. Gen. Genet. 218, 453-459, 1989
A;Title: Nucleotide sequences of Dictyostelium discoideum developmentally regulated cDN
A;Reference number: S05355; MUID:90066348; PMID:2511421
A;Accession: S05356
A;Molecule type: mRNA
A;Residues: 1-317 <SHA>
A;Cross-references: EMBL:X16523; NID:g7321; PIDN:CAA34530.1; PID:g7322

  Query Match             1.2%;  Score 16;  DB 2;  Length 317;
```

```
  Best Local Similarity   100.0%;  Pred. No. 6.3e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db         54 NNNNNNNNNNNNNNNN 69
```

RESULT 17
T02046
prolamin box binding factor - maize
C;Species: Zea mays (maize)
C;Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 29-Oct-1999
C;Accession: T02046
R;Vicente-Carbajosa, J.; Moose, S.P.; Parsons, R.L.; Schmidt, R.J.
Proc. Natl. Acad. Sci. U.S.A. 94, 7685-7690, 1997
A;Title: A maize zinc-finger protein binds the prolamin box in zein gene promoters and i
A;Reference number: Z14511; MUID:97352860; PMID:9207153
A;Accession: T02046
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-328 <VIC>
A;Cross-references: EMBL:U82230; NID:g2393774; PIDN:AAB70119.1; PID:g2393775
C;Genetics:
A;Gene: PBF

```
  Query Match             1.2%;  Score 16;  DB 2;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        310 NNNNNNNNNNNNNNNN 325
```

RESULT 18
T52593
squamosa promoter binding protein homolog 9 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C;Accession: T52593
R;Cardon, G.H.; Hohmann, S.; Nettesheim, K.; Saedler, H.; Huijser, P.
Plant J. 12, 367-377, 1997
A;Title: Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene
A;Reference number: Z24029; MUID:9301089; PMID:9301089
A;Accession: T52593
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-373 <CAR>
A;Cross-references: EMBL:AJ011639; PIDN:CAB56591.1
A;Experimental source: cultivar Landsberg erecta; inflorescence
C;Genetics:
A;Gene: spl9
A;Map position: 2
C;Keywords: DNA binding; transcription factor

```
  Query Match             1.2%;  Score 16;  DB 2;  Length 373;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        262 NNNNNNNNNNNNNNNN 277
```

RESULT 19
T18285
hypothetical protein G2 - slime mold (Dictyostelium discoideum) plasmid
C;Species: Dictyostelium discoideum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C;Accession: T18285; S28720
R;Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh
Genetics 148, 1117-1125, 1998
A;Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1
A;Reference number: Z14684; MUID:98198836; PMID:9539429
A;Accession: T18285
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-401 <RIE>
A;Cross-references: EMBL:U00796; NID:g2702254; PID:g392876; PIDN:AAC18630.1
A;Experimental source: plasmid Ddp5
R;Gurniak, C.B.; Bang, A.G.; Noegel, A.A.
Curr. Genet. 17, 321-325, 1990
A;Title: Transcript and sequence analysis of a 5.1 kb contiguous fragment of Dictyostel
A;Reference number: S28718; MUID:90254839; PMID:2340592
A;Accession: S28720
A;Molecule type: DNA
A;Residues: 155-401 <GUR>
A;Cross-references: EMBL:X53237; NID:g7280; PIDN:CAA37329.1; PID:g7283
A;Experimental source: plasmid Ddp1
A;Note: the authors translated the codon TTT for residue 284 as Glu and GGT for residue
C;Genetics:
A;Gene: d-4
A;Genome: plasmid

```
  Query Match             1.2%;  Score 16;  DB 2;  Length 401;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        743 NNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        358 NNNNNNNNNNNNNNNI 373
```

RESULT 20
A46391
cAMP receptor subtype 3 - slime mold (Dictyostelium sp.)
C;Species: Dictyostelium sp.
C;Date: 18-Oct-1993 #sequence_revision 18-Nov-1994 #text_change 29-Oct-1999
C;Accession: A46391
R;Johnson, R.L.; Saxe III, C.L.; Gollop, R.; Kimmel, A.R.; Devreotes, P.N.
Genes Dev. 7, 273-282, 1993
A;Title: Identification and targeted gene disruption of cAR3, a cAMP receptor subtype e
A;Reference number: A46391; MUID:93170666; PMID:8382181
A;Accession: A46391
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-490 <JOH>
A;Cross-references: GB:S55235; NID:g265733; PIDN:AAB25437.1; PID:g265734
A;Note: sequence extracted from NCBI backbone (NCBIN:125367, NCBIP:125369)
C;Keywords: cAMP binding; G protein-coupled receptor; phosphoprotein; transmembrane pro

```
  Query Match             1.2%;  Score 16;  DB 2;  Length 490;
  Best Local Similarity   100.0%;  Pred. No. 9.2e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        399 NNNNNNNNNNNNNNNN 414
```

RESULT 21
S52830
HMS1 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YM9916.09; protein YMR070w
C;Species: Saccharomyces cerevisiae
C;Date: 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999
C;Accession: S52830; S59820
R;Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, April 1995
A;Reference number: S52814
A;Accession: S52830
A;Molecule type: DNA
A;Residues: 1-490 <PEA>
A;Cross-references: EMBL:Z48952; NID:g763008; PIDN:CAA88795.1; PID:g763017; MIPS:YMR070
A;Experimental source: strain AB972

```
R;Madison, J.; Winston, F.
submitted to the EMBL Data Library, April 1995
A;Reference number: S59820
A;Accession: S59820
A;Molecule type: DNA
A;Residues: 1-490 <MAD>
A;Cross-references: EMBL:U25279; NID:g805129; PIDN:AAC49982.1; PID:g805130
C;Genetics:
A;Gene: SGD:MOT3; HMS1
A;Cross-references: MIPS:YMR070w; SGD:S0004674
A;Map position: 13R

  Query Match             1.2%;  Score 16;  DB 2;  Length 490;
  Best Local Similarity   100.0%;  Pred. No. 9.2e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        743 NNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        143 NNNNNNNNNNNNNNNI 158


RESULT 22
S52641
heat shock transcription factor HSF1 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 01-Aug-1995 #sequence_revision 24-Mar-1999 #text_change 16-Jun-2000
C;Accession: F71447; S52641; S62227; S38873
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A;Reference number: A71400; MUID:98121113; PMID:9461215
A;Accession: F71447
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-495 <BEV>
A;Cross-references: GB:Z97344; NID:g2245126; PIDN:CAB10555.1; PID:g2245134
R;Huebel, A.; Schoeffl, F.
Plant Mol. Biol. 26, 353-362, 1994
A;Title: Arabidopsis heat shock factor: isolation and characterization of the gene and t
A;Reference number: S52641; MUID:95036006; PMID:7948881
A;Accession: S52641
A;Molecule type: DNA
A;Residues: 1-215,'M',217-284,'D',286-335,'T',337-389,'Y',391-469,474-495 <HUE>
A;Cross-references: EMBL:X76167
A;Accession: S62227
A;Molecule type: mRNA
A;Residues: 31-215,'M',217-284,'D',286-335,'T',337-389,'Y',391-469,474-495 <HUW>
R;Huebel, A.; Schoeffl, F.
submitted to the EMBL Data Library, November 1993
A;Description: Arabidopsis heat shock factor: isolation and characterization of the gene
A;Reference number: S38873
A;Accession: S38873
A;Molecule type: DNA
A;Residues: 1-215,'M',217-284,'D',286-335,'T',337-389,'Y',391-461,474-495 <HUF>
A;Cross-references: EMBL:X76167; NID:g429154; PIDN:CAA53761.1; PID:g429155
C;Genetics:
A;Gene: HSF1
A;Map position: 4COP9-4G3845
A;Introns: 111/3
C;Function: <TFC>
A;Description: transcription factor that binds to heat shock promoter elements
C;Superfamily: tomato heat shock transcription factor HSF8; HSF DNA-binding domain homol
C;Keywords: DNA binding; heat shock; leucine zipper; nucleus; stress-induced protein; tr
F;53-148/Domain: HSF DNA-binding domain homology <HSF>
F;181-202/Region: leucine zipper

  Query Match             1.2%;  Score 16;  DB 1;  Length 495;
  Best Local Similarity   100.0%;  Pred. No. 9.2e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        454 NNNNNNNNNNNNNNNN 469


RESULT 23
S53813
RNA helicase - slime mold (Dictyostelium discoideum) (fragment)
C;Species: Dictyostelium discoideum
C;Date: 15-Jul-1995 #sequence_revision 23-Aug-1997 #text_change 02-Feb-2001
C;Accession: S53813; S49262
R;Mahal, B.; Nellen, W.
Biol. Chem. Hoppe-Seyler 375, 759-763, 1994
A;Title: Developmental regulation of DEAD box proteins and cloning of putative RNA heli
A;Reference number: S53813; MUID:95209781; PMID:7695838
A;Accession: S53813
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-566 <MAH>
A;Cross-references: EMBL:X81824; NID:g550330; PIDN:CAA57418.1; PID:g550331
C;Keywords: ATP; nucleotide binding; P-loop
F;189-196/Region: nucleotide-binding motif A (P-loop)
F;298-303/Region: nucleotide-binding motif B
F;302-305/Region: DEAD motif

  Query Match             1.2%;  Score 16;  DB 2;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        743 NNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db         74 NNNNNNNNNNNNNNNI 89


RESULT 24
T18439
hypothetical protein C0380w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2002
C;Accession: T18439
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18439
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-580 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PIDN:CAB11119.1
C;Genetics:
A;Map position: 3
A;Introns: 437/3; 457/2; 482/3; 512/3
A;Note: C0380w

  Query Match             1.2%;  Score 16;  DB 2;  Length 580;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        187 NNNNNNNNNNNNNNNN 202


RESULT 25
T18446
hypothetical protein MAL3P3.11 - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2002
C;Accession: T18446
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
```

```
A;Reference number: Z18935
A;Accession: T18446
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-600 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PIDN:CAB51588.1
C;Genetics:
A;Map position: 3
A;Introns: 437/3; 477/2; 502/3; 532/3
A;Note: MAL3P3.11

  Query Match             1.2%;  Score 16;  DB 2;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        187 NNNNNNNNNNNNNNNN 202

RESULT 26
A71607
Mtn3/RAG1IP-like protein PFB0760w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: A71607
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: A71607
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-686 <GAR>
A;Cross-references: GB:AE001416; GB:AE001362; NID:g3845268; PIDN:AAC71941.1; PID:g384527
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0760w

  Query Match             1.2%;  Score 16;  DB 2;  Length 686;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db         50 NNNNNNNNNNNNNNNN 65

RESULT 27
T51007
hypothetical protein B7J19.60 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T51007
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25286
A;Accession: T51007
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-720 <SCH>
A;Cross-references: EMBL:AL389899; GSPDB:GN00116; NCSP:B7J19.60
A;Experimental source: BAC clone B7J19; strain OR74A
C;Genetics:
A;Gene: NCSP:B7J19.60
A;Map position: 6
A;Introns: 534/2; 692/2

  Query Match             1.2%;  Score 16;  DB 2;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        545 NNNNNNNNNNNNNNNN 560

RESULT 28
S54522
hypothetical protein YMR164c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YM8520.13c
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999
C;Accession: S54522; S54609
R;Hunt, S.; Bowman, S.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54510
A;Accession: S54522
A;Molecule type: DNA
A;Residues: 1-758 <HUN>
A;Cross-references: GB:Z49705; EMBL:Z49700; NID:g825556; PIDN:CAA89800.1; PID:g825569;
A;Experimental source: strain AB972
C;Genetics:
A;Gene: SGD:MSS11
A;Cross-references: SGD:S0004774; MIPS:YMR164c
A;Map position: 13R

  Query Match             1.2%;  Score 16;  DB 2;  Length 758;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        605 NNNNNNNNNNNNNNNN 620

RESULT 29
T08611
hypothetical protein DocA - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08611
R;Aubry, L.; Firtel, R.A.; Iranfar, N.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z16456
A;Accession: T08611
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-831 <AUB>
A;Cross-references: EMBL:AF020409; NID:g2425146; PID:g2425147
A;Experimental source: strain AX4
C;Genetics:
A;Gene: docA

  Query Match             1.2%;  Score 16;  DB 2;  Length 831;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        740 ITNNNNNNNNNNNNNN 755
              ||||||||||||||||
Db        164 ITNNNNNNNNNNNNNN 179

RESULT 30
T08605
hypothetical protein HelE - slime mold (Dictyostelium discoideum) (fragment)
C;Species: Dictyostelium discoideum
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08605
R;Loomis, W.F.; Iranfar, N.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z16451
A;Accession: T08605
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
```

```
A;Residues: 1-947 <LOO>
A;Cross-references: EMBL:AF019981; NID:g2425112; PID:g2425113
A;Experimental source: strain AX4
C;Genetics:
A;Gene: helE

  Query Match             1.2%;  Score 16;  DB 2;  Length 947;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        234 NNNNNNNNNNNNNNNN 249


RESULT 31
S67568
probable membrane protein YDL035c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein D2749
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
C;Accession: S67568
R;Paulin, L.; Saren, A.M.; Laamanen, P.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67560
A;Accession: S67568
A;Molecule type: DNA
A;Residues: 1-961 <PAU>
A;Cross-references: EMBL:Z74083; NID:g1431014; PID:e252988; PID:g1431015; GSPDB:GN00004;
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:GPR1; MIPS:YDL035c
A;Cross-references: SGD:S0002193
A;Map position: 4L
C;Keywords: transmembrane protein
F;57-73/Domain: transmembrane #status predicted <TM1>
F;92-108/Domain: transmembrane #status predicted <TM2>
F;140-156/Domain: transmembrane #status predicted <TM3>
F;181-197/Domain: transmembrane #status predicted <TM4>
F;256-272/Domain: transmembrane #status predicted <TM5>
F;623-639/Domain: transmembrane #status predicted <TM6>
F;662-678/Domain: transmembrane #status predicted <TM7>

  Query Match             1.2%;  Score 16;  DB 2;  Length 961;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        515 NNNNNNNNNNNNNNNN 530


RESULT 32
JC5808
G protein-coupled receptor 1 - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 04-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 19-Apr-2002
C;Accession: JC5808
R;Yun, C.W.; Tamaki, H.; Nakayama, R.; Yamamoto, K.; Kumagai, H.
Biochem. Biophys. Res. Commun. 240, 287-292, 1997
A;Title: G-protein coupled receptor from yeast Saccharomyces cerevisiae.
A;Reference number: JC5808; MUID:98049822; PMID:9388468
A;Accession: JC5808
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-962 <YUN>
C;Comment: This protein monitors the extracellular signal such as nutrition and transduc
C;Genetics:
A;Gene: SGD:GPR1; gpr1
A;Cross-references: SGD:S0002193
F;56-74/Domain: transmembrane #status predicted <TM1>
F;91-109/Domain: transmembrane #status predicted <TM2>
F;139-157/Domain: transmembrane #status predicted <TM3>
F;180-198/Domain: transmembrane #status predicted <TM4>
F;255-273/Domain: transmembrane #status predicted <TM5>
F;622-640/Domain: transmembrane #status predicted <TM6>
F;661-679/Domain: transmembrane #status predicted <TM7>

  Query Match             1.2%;  Score 16;  DB 2;  Length 962;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        516 NNNNNNNNNNNNNNNN 531


RESULT 33
S48404
probable membrane protein YIL130w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 19-Apr-2002
C;Accession: S48404
R;Churcher, C.
submitted to the EMBL Data Library, September 1994
A;Reference number: S48310
A;Accession: S48404
A;Molecule type: DNA
A;Residues: 1-964 <CHU>
A;Cross-references: GB:Z47047; EMBL:Z38059; NID:g603997; PID:g763216; MIPS:YIL130w
C;Genetics:
A;Cross-references: SGD:S0001392
A;Map position: 9L
C;Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluster
C;Keywords: DNA binding; nucleus; transcription regulation; transmembrane protein
F;16-52/Domain: GAL4 zinc binuclear cluster homology <GAL4>
F;289-305/Domain: transmembrane #status predicted <TM1>
F;346-362/Domain: transmembrane #status predicted <TM2>

  Query Match             1.2%;  Score 16;  DB 2;  Length 964;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        853 NNNNNNNNNNNNNNNN 868


RESULT 34
S54067
probable membrane protein YPR042c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YP3085.06c; hypothetical protein YP9499.01c
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 26-Jul-1996 #text_change 19-Apr-2002
C;Accession: S54067; S61063
R;Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54059
A;Accession: S54067
A;Molecule type: DNA
A;Residues: 1-508 <BAD>
A;Cross-references: EMBL:Z49219; NID:g805025; PID:g805026; MIPS:YPR042c
A;Experimental source: strain AB972
R;Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995
A;Reference number: S61058
A;Accession: S61063
A;Molecule type: DNA
A;Residues: 465-1075 <BAW>
A;Cross-references: EMBL:Z68111; MIPS:YPR042c
C;Genetics:
A;Gene: SGD:PUF2
A;Cross-references: SGD:S0006246
A;Map position: 16R
C;Keywords: transmembrane protein
F;711-727/Domain: transmembrane #status predicted <TMM>
```

```
  Query Match             1.2%;  Score 16;  DB 2;  Length 1075;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db       1033 NNNNNNNNNNNNNNNN 1048
```

RESULT 35
T09057
probable protein-histidine kinase (EC 2.7.3.-) DhkC - slime mold (Dictyostelium discoide
N;Alternate names: histidine kinase C
C;Species: Dictyostelium discoideum
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T09057
R;Singleton, C.K.; Mykytka, B.; Zinda, M.
submitted to the EMBL Data Library, October 1997
A;Reference number: Z16542
A;Accession: T09057
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1225 <SIN>
A;Cross-references: EMBL:AF029726; NID:g2605923; PID:g2605924
C;Genetics:
A;Gene: dhkC
A;Map position: 4
C;Keywords: phosphotransferase; protein kinase

```
  Query Match             1.2%;  Score 16;  DB 2;  Length 1225;
  Best Local Similarity   100.0%;  Pred. No. 2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        101 NNNNNNNNNNNNNNNN 116
```

RESULT 36
S14556
asparagine-rich protein (clone 18C1) - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 01-Dec-2000
C;Accession: S14556; S14548
R;Schreiber, L.; Deutsche, U.; Storck, T.; Mueller-Hill, D.
submitted to the EMBL Data Library, December 1989
A;Reference number: S14469
A;Accession: S14556
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1256 <SCH>
A;Cross-references: EMBL:X17485; NID:g9846; PID:g1335715
A;Accession: S14548
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 'Q',210-1251,'NEENDETALDTIIQPPIKKKK' <SC2>
A;Cross-references: EMBL:X17488; NID:g9852; PID:g9853

```
  Query Match             1.2%;  Score 16;  DB 2;  Length 1256;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db       1078 NNNNNNNNNNNNNNNN 1093
```

RESULT 37
T08607
hypothetical protein DG2033 - slime mold (Dictyostelium discoideum) (fragment)
C;Species: Dictyostelium discoideum
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08607
R;Loomis, W.F.; Iranfar, N.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z16451
A;Accession: T08607
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1271 <LOO>
A;Cross-references: EMBL:AF020282; NID:g2425132; PID:g2425133
A;Experimental source: strain AX3
C;Genetics:
A;Gene: DG2033

```
  Query Match             1.2%;  Score 16;  DB 2;  Length 1271;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db       1047 NNNNNNNNNNNNNNNN 1062
```

RESULT 38
T18288
ABC transport protein - slime mold (Dictyostelium discoideum) (fragment)
C;Species: Dictyostelium discoideum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18288
R;Loomis, W.F.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z18857
A;Accession: T18288
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1336 <LOO>
A;Cross-references: EMBL:U66526; NID:g1513297; PID:g1513298; PIDN:AAB06789.1
C;Genetics:
A;Gene: abcA

```
  Query Match             1.2%;  Score 16;  DB 2;  Length 1336;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        789 NNNNNNNNNNNNNNNN 804
```

RESULT 39
S14871
suppressor two of zeste protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-2000
C;Accession: S14871; S16845
R;Brunk, B.P.; Adler, P.N.
submitted to the EMBL Data Library, November 1990
A;Description: The Drosophila regulatory gene suppressor two of zeste encodes a large z
A;Reference number: S14871
A;Accession: S14871
A;Molecule type: mRNA
A;Residues: 1-1365 <BRU>
A;Cross-references: EMBL:X56799; NID:g8528; PID:g8529
R;Brunk, B.P.; Adler, P.N.
Nucleic Acids Res. 19, 3149, 1991
A;Title: The sequence of the Drosophila regulatory gene Suppressor two of zeste.
A;Reference number: S16845; MUID:91279476; PMID:2057369
A;Accession: S16845
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-612,614-784,'N',786-830,'R',832-965,967-1064,'E',1066-1096,'N',1097-1286,
A;Cross-references: EMBL:X56798; NID:g8526; PIDN:CAA40134.1; PID:g8527
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1990
C;Genetics:

```
A;Gene: FlyBase:Su(z)2
A;Cross-references: FlyBase:FBgn0008654
A;Introns: 55/1; 139/2; 219/3
C;Superfamily: Drosophila suppressor protein of zeste; RING finger homology
C;Keywords: DNA binding; nucleus; zinc finger
F;31-79/Domain: RING finger homology <RNG>

  Query Match             1.2%;  Score 16;  DB 2;  Length 1365;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db       1077 NNNNNNNNNNNNNNNN 1092


RESULT 40
T14004
trfA protein - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C;Accession: T14004
R;Saito, J.; Adachi, H.; Sutoh, K.
J. Biol. Chem. 273, 24654-24659, 1998
A;Title: Dictyostelium TRFA homologous to yeast Ssn6 is required for normal growth and e
A;Reference number: Z17852; MUID:98406112; PMID:9733762
A;Accession: T14004
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1390 <SAI>
A;Cross-references: EMBL:AB009080; NID:d1228566; PID:d1034109; PIDN:BAA33143.1
C;Genetics:
A;Gene: trfA
A;Introns: 333/3; 364/3; 637/1

  Query Match             1.2%;  Score 16;  DB 2;  Length 1390;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        830 NNNNNNNNNNNNNNNN 845


Search completed: October  2, 2003, 16:36:14
Job time : 78 secs
```

OM protein - protein search, using sw model

Run on: October 2, 2003, 16:27:41 ; Search time 29 Seconds
(without alignments)
2088.634 Million cell updates/sec

Title: US-10-006-780-2
Perfect score: 1288
Sequence: 1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	17	1.3	666	1	YEA7_YEAST	P40002 saccharomyc
2	17	1.3	758	1	CC27_YEAST	P38042 saccharomyc
3	17	1.3	1093	1	PI4K_DICDI	P54677 dictyosteli
4	17	1.3	1178	1	PH81_YEAST	P17442 saccharomyc
5	16	1.2	317	1	AAC4_DICDI	P14198 dictyosteli
6	16	1.2	451	1	MYBH_DICDI	P34127 dictyosteli
7	16	1.2	490	1	CAR3_DICDI	P35352 dictyosteli
8	16	1.2	490	1	MOT3_YEAST	P54785 saccharomyc
9	16	1.2	495	1	HSF1_ARATH	P41151 arabidopsis
10	16	1.2	640	1	RAEP_CANAL	O93831 candida alb
11	16	1.2	732	1	KMHB_DICDI	P90648 dictyosteli
12	16	1.2	735	1	CIGB_DICDI	Q94481 dictyosteli
13	16	1.2	758	1	YM38_YEAST	Q03825 saccharomyc
14	16	1.2	961	1	GPR1_YEAST	Q12361 saccharomyc
15	16	1.2	964	1	YIN0_YEAST	P40467 saccharomyc
16	16	1.2	989	1	PTP3_DICDI	P54637 dictyosteli
17	16	1.2	1365	1	SUZ2_DROME	P25172 drosophila
18	16	1.2	1584	1	KYK1_DICDI	P18160 dictyosteli
19	16	1.2	1585	1	P3K3_DICDI	P54675 dictyosteli
20	16	1.2	1743	1	TAGC_DICDI	Q23868 dictyosteli
21	16	1.2	1858	1	P3K2_DICDI	P54674 dictyosteli
22	16	1.2	1905	1	TAGB_DICDI	P54683 dictyosteli
23	15	1.2	451	1	ARP2_PLAFA	P13824 plasmodium
24	15	1.2	612	1	YP68_YEAST	Q08925 saccharomyc
25	15	1.2	837	1	CCR4_YEAST	P31384 saccharomyc
26	15	1.2	888	1	YGB4_YEAST	P25339 saccharomyc
27	15	1.2	2278	1	FAB1_YEAST	P34756 saccharomyc
28	14	1.1	643	1	CTK2_XENLA	P79955 xenopus lae
29	14	1.1	700	1	NCD_DROME	P20480 drosophila
30	14	1.1	730	1	KCM1_XENLA	Q91636 xenopus lae
31	14	1.1	745	1	ATK2_ARATH	P46864 arabidopsis
32	14	1.1	754	1	ATK3_ARATH	P46875 arabidopsis
33	14	1.1	793	1	ATK1_ARATH	Q07970 arabidopsis
34	14	1.1	832	1	KLP1_SCHPO	Q92376 schizosacch
35	14	1.1	858	1	CYAG_DICDI	Q03101 dictyosteli
36	14	1.1	955	1	KINL_LEICH	P46865 leishmania
37	14	1.1	1770	1	RI15_YEAST	P43565 saccharomyc
38	13	1.0	201	1	YM8D_YEAST	P38429 saccharomyc
39	13	1.0	426	1	HKLB_LYCES	O22300 lycopersico
40	13	1.0	554	1	KLP3_SCHPO	Q9us60 schizosacch
41	13	1.0	669	1	FPS1_YEAST	P23900 saccharomyc
42	13	1.0	708	1	GBF_DICDI	P36417 dictyosteli
43	13	1.0	851	1	NUD1_YEAST	P32336 saccharomyc
44	13	1.0	1151	1	GRR1_YEAST	P24814 saccharomyc
45	12	0.9	212	1	YMF3_YEAST	Q04978 saccharomyc
46	12	0.9	537	1	ARP_PLAFA	P04931 plasmodium
47	12	0.9	590	1	HMDF_DROME	P07548 drosophila
48	12	0.9	624	1	DSK1_CYLFU	Q39493 cylindrothe
49	12	0.9	647	1	KNRL_DROME	P13054 drosophila
50	12	0.9	671	1	KF2C_MACFA	Q951p1 macaca fasc
51	12	0.9	671	1	KF2C_RAT	Q62909 rattus norv
52	12	0.9	679	1	KIF2_HUMAN	O00139 homo sapien
53	12	0.9	682	1	KIF2_XENLA	Q91637 xenopus lae
54	12	0.9	716	1	KIF2_MOUSE	P28740 mus musculu
55	12	0.9	718	1	KF2C_CRIGR	P70096 cricetulus
56	12	0.9	721	1	KF2C_MOUSE	Q922s8 mus musculu
57	12	0.9	725	1	KF2C_HUMAN	Q99661 homo sapien
58	12	0.9	727	1	BRC1_DROME	Q01295 drosophila
59	12	0.9	914	1	AZF1_YEAST	P41696 saccharomyc
60	12	0.9	1177	1	SP97_DICDI	Q95zg3 dictyosteli
61	12	0.9	1341	1	YL78_YEAST	Q05854 saccharomyc
62	12	0.9	1570	1	P3K1_DICDI	P54673 dictyosteli
63	12	0.9	2452	1	RPB1_PLAFD	P14248 plasmodium
64	11	0.9	172	1	PKIA_DICDI	Q23921 dictyosteli
65	11	0.9	331	1	COL6_ARATH	Q9lqz7 arabidopsis
66	11	0.9	443	1	CAR4_DICDI	Q9tx43 dictyosteli
67	11	0.9	448	1	AAC2_DICDI	P14196 dictyosteli
68	11	0.9	472	1	HMCA_DROME	P09085 drosophila
69	11	0.9	672	1	OSM3_CAEEL	P46873 caenorhabdi
70	11	0.9	673	1	KFC1_HUMAN	Q9bw19 homo sapien
71	11	0.9	676	1	RLM1_YEAST	Q12224 saccharomyc
72	11	0.9	729	1	KAR3_YEAST	P17119 saccharomyc
73	11	0.9	784	1	KLP6_SCHPO	O59751 schizosacch
74	11	0.9	793	1	REGA_DICDI	Q23917 dictyosteli
75	11	0.9	805	1	KIP3_YEAST	P53086 saccharomyc
76	11	0.9	935	1	KINH_SYNRA	O43093 syncephalas
77	11	0.9	987	1	ATK4_ARATH	O81635 arabidopsis
78	11	0.9	1225	1	KF4A_CHICK	Q90640 gallus gall
79	11	0.9	1226	1	KF4A_XENLA	Q91784 xenopus lae
80	11	0.9	1230	1	ST20_CANAL	Q92212 candida alb
81	11	0.9	1231	1	KF4A_MOUSE	P33174 mus musculu
82	11	0.9	1232	1	KF4A_HUMAN	O95239 homo sapien
83	11	0.9	2339	1	RPC1_PLAFA	P27625 plasmodium
84	10	0.8	240	1	MNTB_LISIN	Q92af9 listeria in
85	10	0.8	240	1	MNTB_LISMO	Q8y651 listeria mo
86	10	0.8	368	1	YBI1_YEAST	P38180 saccharomyc
87	10	0.8	372	1	COLB_ARATH	Q9sse5 arabidopsis
88	10	0.8	522	1	IL9R_HUMAN	Q01113 homo sapien
89	10	0.8	525	1	ESR1_MICUN	P57753 micropogoni
90	10	0.8	623	1	PNT1_DROME	P51022 drosophila
91	10	0.8	648	1	KAPC_DICDI	P34099 dictyosteli
92	10	0.8	718	1	PNT2_DROME	P51023 drosophila
93	10	0.8	720	1	ABI3_ARATH	Q01593 arabidopsis
94	10	0.8	726	1	NFI1_YEAST	Q12216 saccharomyc
95	10	0.8	823	1	SCH9_YEAST	P11792 saccharomyc
96	10	0.8	865	1	E78A_DROME	P45447 drosophila
97	10	0.8	972	1	OMB_DROME	Q24432 drosophila
98	10	0.8	1038	1	KF17_MOUSE	Q99pw8 mus musculu
99	10	0.8	1080	1	HDC_DROME	Q9n2m8 drosophila
100	10	0.8	1195	1	YK76_YEAST	P36168 saccharomyc
101	10	0.8	1228	1	ATC_PLAFK	Q08853 plasmodium
102	10	0.8	1419	1	MDR_PLAFF	P13568 plasmodium
103	10	0.8	1584	1	U104_CAEEL	P23678 caenorhabdi
104	10	0.8	1648	1	KF14_HUMAN	Q15058 homo sapien
105	10	0.8	1749	1	K13A_MOUSE	Q9eqw7 mus musculu
106	10	0.8	1805	1	K13A_HUMAN	Q9h1h9 homo sapien

107	10	0.8	1826	1	K13B_HUMAN	Q9nqt8	homo sapien
108	9	0.7	375	1	CAR2_DICDI	P34907	dictyosteli
109	9	0.7	377	1	PTP2_DICDI	P34138	dictyosteli
110	9	0.7	384	1	KF25_HUMAN	Q9ui14	homo sapien
111	9	0.7	398	1	HKL1_ARATH	P46639	arabidopsis
112	9	0.7	403	1	GBA8_DICDI	P34046	dictyosteli
113	9	0.7	416	1	PKD1_DICDI	P34100	dictyosteli
114	9	0.7	472	1	NU49_YEAST	Q02199	saccharomyc
115	9	0.7	473	1	DCOR_CANAL	P78599	candida alb
116	9	0.7	508	1	EGR1_RAT	P08154	rattus norv
117	9	0.7	528	1	CTK1_YEAST	Q03957	saccharomyc
118	9	0.7	564	1	BMRP_CANAL	P28873	candida alb
119	9	0.7	577	1	P2B1_DROME	P48456	drosophila
120	9	0.7	578	1	PSP2_YEAST	P50109	saccharomyc
121	9	0.7	598	1	KLP3_CAEEL	P45962	caenorhabdi
122	9	0.7	658	1	UVRB_HELPJ	Q9zka0	helicobacte
123	9	0.7	658	1	UVRB_HELPY	P94846	helicobacte
124	9	0.7	675	1	UVRB_NEIGO	Q50939	neisseria g
125	9	0.7	675	1	UVRB_NEIMA	P56996	neisseria m
126	9	0.7	675	1	UVRB_NEIMB	O33395	neisseria m
127	9	0.7	682	1	CONN_DROME	Q01819	drosophila
128	9	0.7	696	1	UVRB_RALSO	Q8y0n2	ralstonia s
129	9	0.7	713	1	YHC4_YEAST	P38741	saccharomyc
130	9	0.7	719	1	NRP1_YEAST	P32770	saccharomyc
131	9	0.7	740	1	KRE6_CANAL	P87023	candida alb
132	9	0.7	749	1	MAD1_YEAST	P40957	saccharomyc
133	9	0.7	770	1	KLPA_EMENI	P28739	emericella
134	9	0.7	779	1	SRP_DROME	P52172	drosophila
135	9	0.7	782	1	SP4_MOUSE	Q62445	mus musculu
136	9	0.7	784	1	SP4_HUMAN	Q02446	homo sapien
137	9	0.7	868	1	NRG2_RAT	O35569	rattus norv
138	9	0.7	904	1	ABRU_DROME	Q24174	drosophila
139	9	0.7	928	1	PTP3_YEAST	P40048	saccharomyc
140	9	0.7	960	1	DLG1_DROME	P31007	drosophila
141	9	0.7	1000	1	S155_YEAST	P43612	saccharomyc
142	9	0.7	1091	1	JSN1_YEAST	P47135	saccharomyc
143	9	0.7	1094	1	YB00_YEAST	P38114	saccharomyc
144	9	0.7	1188	1	OSH1_YEAST	P35845	saccharomyc
145	9	0.7	1227	1	JIP3_DROME	Q9gqf1	drosophila
146	9	0.7	1398	1	TOP2_PLAFK	P41001	plasmodium
147	9	0.7	1407	1	CYAA_DICDI	Q03100	dictyosteli
148	9	0.7	1556	1	PROS_DROVI	Q9u6a1	drosophila
149	9	0.7	1677	1	VIT_ACITR	Q90243	acipenser t
150	9	0.7	3092	1	IRA1_YEAST	P18963	saccharomyc
151	8	0.6	115	1	SER1_GALME	O96614	galleria me
152	8	0.6	117	1	DHA_CRAPL	P22238	craterostig
153	8	0.6	121	1	RS6_RICCN	Q92jk3	rickettsia
154	8	0.6	121	1	RS6_RICPR	Q9zea6	rickettsia
155	8	0.6	150	1	ASNC_HAEIN	P44337	haemophilus
156	8	0.6	193	1	UBC6_HUMAN	P51965	homo sapien
157	8	0.6	193	1	UBC6_MOUSE	P52482	mus musculu
158	8	0.6	215	1	FGFA_RAT	P70492	rattus norv
159	8	0.6	220	1	SER2_GALME	O96615	galleria me
160	8	0.6	260	1	DH10_ARATH	P42759	arabidopsis
161	8	0.6	265	1	DH47_ARATH	P31168	arabidopsis
162	8	0.6	265	1	JOIN_LYCES	Q9fuy6	lycopersico
163	8	0.6	284	1	SMX5_SCHMA	Q26604	schistosoma
164	8	0.6	288	1	KADC_SOLTU	Q8hsw1	solanum tub
165	8	0.6	298	1	MYOD_CHICK	P16075	gallus gall
166	8	0.6	329	1	CAHX_FLAPR	P46281	flaveria pr
167	8	0.6	330	1	CAH1_FLALI	P46512	flaveria li
168	8	0.6	330	1	CAHX_FLABR	P46511	flaveria br
169	8	0.6	331	1	CAHX_FLABI	P46510	flaveria bi
170	8	0.6	334	1	YPD6_CAEEL	P55112	caenorhabdi
171	8	0.6	339	1	RMAR_CANGA	P21358	candida gla
172	8	0.6	354	1	URE2_YEAST	P23202	saccharomyc
173	8	0.6	361	1	CGD2_ARATH	P42752	arabidopsis
174	8	0.6	367	1	CSP_PLAYO	P06914	plasmodium
175	8	0.6	369	1	DAG2_ARATH	Q9zpy0	arabidopsis
176	8	0.6	372	1	B4G2_HUMAN	O60909	h beta-1,4-
177	8	0.6	372	1	MSSP_HUMAN	P29558	homo sapien
178	8	0.6	396	1	RMAR_YEAST	P02381	saccharomyc
179	8	0.6	397	1	ODPA_PEA	P52902	pisum sativ
180	8	0.6	397	1	YMP7_YEAST	Q04359	saccharomyc
181	8	0.6	413	1	E2F4_HUMAN	Q16254	homo sapien
182	8	0.6	429	1	HML2_HELRO	Q01622	helobdella
183	8	0.6	437	1	TIS1_DROME	P47980	drosophila
184	8	0.6	443	1	YEW0_YEAST	P39959	saccharomyc
185	8	0.6	446	1	CHMO_BETVU	O22553	beta vulgar
186	8	0.6	455	1	YB12_YEAST	P38288	saccharomyc
187	8	0.6	461	1	US45_LACLC	P22865	lactococcus
188	8	0.6	516	1	P54_ENTFC	P13692	enterococcu
189	8	0.6	535	1	C4C3_DROME	Q9va27	drosophila
190	8	0.6	554	1	HAP4_YEAST	P14064	saccharomyc
191	8	0.6	567	1	NU2M_HANWI	P48906	hansenula w
192	8	0.6	569	1	MANC_PIRSP	P55298	piromyces s
193	8	0.6	571	1	MANB_PIRSP	P55297	piromyces s
194	8	0.6	583	1	NIR_BETVE	P38500	betula verr
195	8	0.6	586	1	RT67_ECOLI	P21325	escherichia
196	8	0.6	604	1	SP20_YEAST	P50875	saccharomyc
197	8	0.6	612	1	TIPD_DICDI	O15736	dictyosteli
198	8	0.6	653	1	YJU4_YEAST	P39531	saccharomyc
199	8	0.6	663	1	DUS8_MOUSE	O09112	mus musculu
200	8	0.6	669	1	UVRB_XYLFA	Q9per1	xylella fas
201	8	0.6	670	1	UVRB_METAC	Q8tks3	methanosarc
202	8	0.6	670	1	UVRB_METMA	Q8prz9	methanosarc
203	8	0.6	673	1	UVRB_XANAC	Q8pjb1	xanthomonas
204	8	0.6	673	1	UVRB_XANCP	Q8p7x1	xanthomonas
205	8	0.6	675	1	ECR_AEDAE	P49880	aedes aegyp
206	8	0.6	697	1	SIM_DROME	P05709	drosophila
207	8	0.6	717	1	ARA_DROME	Q24248	drosophila
208	8	0.6	784	1	KL68_DROME	P46867	drosophila
209	8	0.6	786	1	STUB_DROME	Q05319	drosophila
210	8	0.6	834	1	MPT5_YEAST	P39016	saccharomyc
211	8	0.6	837	1	UBPA_DICDI	P54201	dictyosteli
212	8	0.6	841	1	PSP1_YEAST	P50896	saccharomyc
213	8	0.6	847	1	R132_ARATH	Q9ste5	arabidopsis
214	8	0.6	865	1	LAC9_KLULA	P08657	kluyveromyc
215	8	0.6	870	1	CLPC_CHLMU	Q9pka8	chlamydia m
216	8	0.6	880	1	TRK2_SCHPO	Q10065	schizosacch
217	8	0.6	883	1	KLP5_SCHPO	O14343	schizosacch
218	8	0.6	959	1	G2D1_HUMAN	Q9uhl9	h general t
219	8	0.6	962	1	GUNA_PSEFL	P10476	pseudomonas
220	8	0.6	970	1	DA81_YEAST	P21657	saccharomyc
221	8	0.6	970	1	Y277_MYCGE	Q49409	mycoplasma
222	8	0.6	975	1	CRM_DROSE	Q8mx88	drosophila
223	8	0.6	976	1	MGR_DROME	P91685	drosophila
224	8	0.6	1028	1	OVO_DROME	P51521	drosophila
225	8	0.6	1043	1	FTFI_DROME	P33244	drosophila
226	8	0.6	1067	1	SGG_DROME	P18431	drosophila
227	8	0.6	1075	1	Y124_METJA	Q57588	methanococc
228	8	0.6	1099	1	PLC1_CANAL	O13433	candida alb
229	8	0.6	1104	1	G2D1_MOUSE	Q9ji57	mus musculu
230	8	0.6	1242	1	IRS1_HUMAN	P35568	homo sapien
231	8	0.6	1314	1	SWI1_YEAST	P09547	saccharomyc
232	8	0.6	1321	1	IRS2_MOUSE	P81122	mus musculu
233	8	0.6	1324	1	IRS2_HUMAN	Q9y4h2	homo sapien
234	8	0.6	1403	1	PROS_DROME	P29617	drosophila
235	8	0.6	1468	1	N153_RAT	P49791	rattus norv
236	8	0.6	1475	1	N153_HUMAN	P49790	homo sapien
237	8	0.6	1529	1	PDRF_YEAST	Q04182	saccharomyc
238	8	0.6	1597	1	SOL_DROME	P27398	drosophila
239	8	0.6	1679	1	FUR2_DROME	P30432	drosophila
240	8	0.6	1704	1	VIT1_FUNHE	Q90508	fundulus he
241	8	0.6	1850	1	VIT2_CHICK	P02845	gallus gall
242	8	0.6	2175	1	HMCU_DROME	P10180	drosophila
243	8	0.6	3005	1	ZFH2_DROME	P28167	drosophila
244	8	0.6	3726	1	TRX_DROME	P20659	drosophila
245	7	0.5	46	1	VIT_ANAPL	P56530	anas platyr
246	7	0.5	70	1	YNCH_ECOLI	P76118	escherichia
247	7	0.5	91	1	MINE_WIGBR	Q8d2i1	wiggleswort
248	7	0.5	92	1	HMGC_TETTH	P40627	tetrahymena
249	7	0.5	97	1	YNI3_METTL	P25769	methanococc
250	7	0.5	98	1	HMEN_HELTR	P23397	helobdella
251	7	0.5	99	1	YO10_BPHP1	P51712	bacteriopha
252	7	0.5	100	1	LG1_TETTH	P11873	tetrahymena

253	7	0.5	102	1	108_LYCES	Q43495 lycopersico
254	7	0.5	102	1	MIH_PENVA	P55322 penaeus van
255	7	0.5	104	1	YJO4_YEAST	P47009 saccharomyc
256	7	0.5	112	1	VE4_HPV29	P51897 human papil
257	7	0.5	115	1	CTE9_MOUSE	Q9cr37 mus musculu
258	7	0.5	117	1	YFB9_YEAST	P43576 saccharomyc
259	7	0.5	119	1	M126_CHICK	P28318 gallus gall
260	7	0.5	122	1	PA25_TRIST	P82896 trimeresuru
261	7	0.5	123	1	RL7_WIGBR	Q8d234 wiggleswort
262	7	0.5	124	1	YBF3_YEAST	P38190 saccharomyc
263	7	0.5	128	1	DHX1_ARATH	P25863 arabidopsis
264	7	0.5	130	1	Y816_ARATH	Q818q8 arabidopsis
265	7	0.5	139	1	DH1_HORVU	P12951 hordeum vul
266	7	0.5	143	1	DH2_HORVU	P12952 hordeum vul
267	7	0.5	145	1	DH11_GOSHI	P09442 gossypium h
268	7	0.5	148	1	PFD2_ARATH	Q91j98 arabidopsis
269	7	0.5	149	1	DH15_WHEAT	Q00742 triticum ae
270	7	0.5	150	1	ZFP2_ARATH	Q39261 arabidopsis
271	7	0.5	151	1	DH1D_ORYSA	P22913 oryza sativ
272	7	0.5	155	1	DHB_CRAPL	P22239 craterostig
273	7	0.5	160	1	KF1C_MOUSE	O35071 mus musculu
274	7	0.5	160	1	NUSB_CHLPN	Q9z6s0 chlamydia p
275	7	0.5	161	1	DH3_HORVU	P12948 hordeum vul
276	7	0.5	163	1	DH21_ORYSA	P12253 oryza sativ
277	7	0.5	164	1	DH1B_ORYSA	P22911 oryza sativ
278	7	0.5	164	1	DH1C_ORYSA	P22912 oryza sativ
279	7	0.5	166	1	YEF4_YEAST	P32617 saccharomyc
280	7	0.5	168	1	DH1_MAIZE	P12950 zea mays (m
281	7	0.5	168	1	PIG1_DROME	P26023 drosophila
282	7	0.5	169	1	Y358_BUCAI	P57439 buchnera ap
283	7	0.5	170	1	COPT_ARATH	Q39065 arabidopsis
284	7	0.5	170	1	DEF_AGRT5	Q8uid1 agrobacteri
285	7	0.5	170	1	HSB7_HUMAN	Q9uby9 homo sapien
286	7	0.5	174	1	DEF_RHIME	Q92sh6 rhizobium m
287	7	0.5	184	1	DHLE_RAPSA	P21298 raphanus sa
288	7	0.5	185	1	DH14_ARATH	P42763 arabidopsis
289	7	0.5	188	1	ERBP_RAT	P06911 rattus norv
290	7	0.5	190	1	ARF_GIALA	P26991 giardia lam
291	7	0.5	193	1	LB12_ARATH	Q8lbw3 arabidopsis
292	7	0.5	194	1	YK42_PYRFU	Q8tzf3 pyrococcus
293	7	0.5	198	1	AMOS_DROME	Q9y0a7 drosophila
294	7	0.5	198	1	C4_GIALA	Q01832 giardia lam
295	7	0.5	198	1	CAS1_CAVPO	P04656 cavia porce
296	7	0.5	200	1	HMGH_STRPU	P40644 strongyloce
297	7	0.5	203	1	YG3R_YEAST	P53288 saccharomyc
298	7	0.5	207	1	NU6M_HANWI	P48927 hansenula w
299	7	0.5	207	1	YJY2_YEAST	P47087 saccharomyc
300	7	0.5	211	1	SOC1_HUMAN	O15524 homo sapien
301	7	0.5	212	1	SOC1_MOUSE	O35716 mus musculu
302	7	0.5	212	1	SOC1_RAT	Q9qx78 rattus norv
303	7	0.5	213	1	HP1_DROVI	P29227 drosophila
304	7	0.5	216	1	ANSP_EIMTE	P15744 eimeria ten
305	7	0.5	218	1	ESM2_DROME	O97177 drosophila
306	7	0.5	218	1	WR12_ARATH	Q93wy4 arabidopsis
307	7	0.5	219	1	COA4_TTV1	P19273 thermoprote
308	7	0.5	222	1	COLI_THUOB	Q9ygk2 t corticotr
309	7	0.5	223	1	DEOC_STRP3	Q8nzg9 streptococc
310	7	0.5	223	1	DEOC_STRPY	Q99y51 streptococc
311	7	0.5	224	1	RM01_RECAM	O21235 reclinomona
312	7	0.5	225	1	DH4_HORVU	P12949 hordeum vul
313	7	0.5	228	1	DH25_ORYSA	P30287 oryza sativ
314	7	0.5	229	1	COAT_TAV	P23627 tomato aspe
315	7	0.5	229	1	LB11_ARATH	Q9sk08 arabidopsis
316	7	0.5	230	1	DAG_ANTMA	Q38732 antirrhinum
317	7	0.5	235	1	YHE4_YEAST	P38727 saccharomyc
318	7	0.5	238	1	CW14_YEAST	O13547 saccharomyc
319	7	0.5	238	1	MNBA_MAIZE	P38564 zea mays (m
320	7	0.5	242	1	NAGB_MYCPE	Q8ewm7 mycoplasma
321	7	0.5	243	1	ERF2_ARATH	O80338 arabidopsis
322	7	0.5	243	1	IM17_ARATH	Q9sp35 arabidopsis
323	7	0.5	244	1	VC23_VACCC	P21090 vaccinia vi
324	7	0.5	244	1	YENR_YEREN	P54295 yersinia en
325	7	0.5	246	1	AXIB_ARATH	Q38829 arabidopsis

326	7	0.5	247	1	DB83_HUMAN	P57088 homo sapien
327	7	0.5	247	1	DB83_MOUSE	Q9cr67 mus musculu
328	7	0.5	247	1	DB83_RAT	Q9z142 rattus norv
329	7	0.5	250	1	UBC6_YEAST	P33296 saccharomyc
330	7	0.5	251	1	YC52_METJA	Q58648 methanococc
331	7	0.5	253	1	LI39_CAEEL	P34684 caenorhabdi
332	7	0.5	254	1	EMD_HUMAN	P50402 homo sapien
333	7	0.5	254	1	S5A2_MOUSE	Q99n99 mus musculu
334	7	0.5	255	1	BDNF_CAVPO	O70183 cavia porce
335	7	0.5	255	1	HXD4_HUMAN	P09016 homo sapien
336	7	0.5	258	1	VC23_VACCL	P19063 vaccinia vi
337	7	0.5	259	1	YG31_YEAST	P53269 saccharomyc
338	7	0.5	260	1	OLG1_MOUSE	Q9jkn5 mus musculu
339	7	0.5	260	1	ZFP4_ARATH	Q39263 arabidopsis
340	7	0.5	262	1	COR4_WHEAT	P46524 triticum ae
341	7	0.5	263	1	COLI_RANCA	P11885 rana catesb
342	7	0.5	263	1	PSP1_ARATH	Q42029 arabidopsis
343	7	0.5	265	1	PSP2_ARATH	O49344 arabidopsis
344	7	0.5	265	1	YIT6_YEAST	P40572 saccharomyc
345	7	0.5	266	1	SUE1_PYRAE	Q8zu79 pyrobaculum
346	7	0.5	268	1	LB13_ARATH	Q9at61 arabidopsis
347	7	0.5	268	1	UPK_BUCBP	P59523 buchnera ap
348	7	0.5	273	1	FWDC_METJA	Q58571 methanococc
349	7	0.5	274	1	PSBS_SPIOL	Q02060 spinacia ol
350	7	0.5	275	1	HXA5_HETFR	Q9ia23 heterodontu
351	7	0.5	280	1	VIE3_HCMVT	P06434 human cytom
352	7	0.5	284	1	CAS1_RAT	P02661 rattus norv
353	7	0.5	285	1	Y007_BORBU	O51040 borrelia bu
354	7	0.5	285	1	YKG2_CAEEL	P46552 caenorhabdi
355	7	0.5	292	1	YO42_CAEEL	P34680 caenorhabdi
356	7	0.5	297	1	TX20_MOUSE	Q9es03 mus musculu
357	7	0.5	304	1	WR13_ARATH	Q9svb7 arabidopsis
358	7	0.5	306	1	EC40_DAUCA	Q07322 daucus caro
359	7	0.5	308	1	ZDH7_HUMAN	Q9nxf8 homo sapien
360	7	0.5	310	1	UDP_HUMAN	Q16831 homo sapien
361	7	0.5	313	1	TFS2_DROME	P20232 drosophila
362	7	0.5	313	1	YM58_YEAST	Q03695 saccharomyc
363	7	0.5	314	1	SINA_DROME	P21461 drosophila
364	7	0.5	314	1	SINA_DROVI	P29304 drosophila
365	7	0.5	323	1	OTX1_BRARE	Q91994 brachydanio
366	7	0.5	324	1	YL70_ARCFU	O28112 archaeoglob
367	7	0.5	325	1	YHU3_YEAST	P38844 saccharomyc
368	7	0.5	328	1	CAHC_PEA	P17067 pisum sativ
369	7	0.5	328	1	DLX2_HUMAN	Q07687 homo sapien
370	7	0.5	329	1	PSTB_MYCGE	P47650 mycoplasma
371	7	0.5	333	1	PIX2_CHICK	O93385 gallus gall
372	7	0.5	334	1	LYTE_BACSU	P54421 bacillus su
373	7	0.5	335	1	HXDD_HUMAN	P35453 homo sapien
374	7	0.5	336	1	CAHC_ARATH	P27140 arabidopsis
375	7	0.5	338	1	FEN_METAC	Q8tiy5 methanosarc
376	7	0.5	338	1	FEN_METMA	Q8pyf6 methanosarc
377	7	0.5	338	1	HRCA_THEMA	Q9wzv5 thermotoga
378	7	0.5	338	1	OTX3_BRARE	Q90267 brachydanio
379	7	0.5	339	1	HXDD_MOUSE	P70217 mus musculu
380	7	0.5	339	1	TRPD_BRUME	Q8yhf7 brucella me
381	7	0.5	344	1	YB1E_SCHPO	P87179 schizosacch
382	7	0.5	347	1	E13A_SOYBN	Q03773 glycine max
383	7	0.5	347	1	UTR2_YEAST	P32623 saccharomyc
384	7	0.5	347	1	VIT3_CHICK	Q91025 gallus gall
385	7	0.5	352	1	STSY_CATRO	P18417 catharanthu
386	7	0.5	355	1	HKL6_LYCES	O22299 lycopersico
387	7	0.5	355	1	YDP6_SCHPO	O14009 schizosacch
388	7	0.5	357	1	V136_DICDI	Q94480 dictyosteli
389	7	0.5	363	1	CYA2_RHIME	Q52915 rhizobium m
390	7	0.5	364	1	NK61_MESAU	Q60554 mesocricetu
391	7	0.5	365	1	NK61_MOUSE	Q99ma9 mus musculu
392	7	0.5	365	1	NK61_RAT	O35762 rattus norv
393	7	0.5	366	1	GGPP_SINAL	Q43133 sinapis alb
394	7	0.5	367	1	NK61_HUMAN	P78426 homo sapien
395	7	0.5	367	1	Y797_METJA	Q58207 methanococc
396	7	0.5	368	1	ROX1_YEAST	P25042 saccharomyc
397	7	0.5	369	1	MAGA_HUMAN	P43363 homo sapien
398	7	0.5	372	1	COQ2_YEAST	P32378 saccharomyc

399	7	0.5	375	1	OTC_PEA	Q43814 pisum sativ
400	7	0.5	376	1	YBC9_YEAST	P38201 saccharomyc
401	7	0.5	377	1	VE2_HPV13	Q02263 human papil
402	7	0.5	377	1	VMAT_PI2HT	P24266 human parai
403	7	0.5	377	1	Y835_METJA	Q58245 methanococc
404	7	0.5	378	1	KLP2_BOMMO	P46874 bombyx mori
405	7	0.5	379	1	HMB1_SOYBN	P46608 glycine max
406	7	0.5	382	1	TP6A_PYRAB	Q9v134 pyrococcus
407	7	0.5	382	1	TP6A_PYRFU	Q8u0k9 pyrococcus
408	7	0.5	382	1	TP6A_PYRHO	O59209 pyrococcus
409	7	0.5	383	1	GIC2_YEAST	Q06648 saccharomyc
410	7	0.5	383	1	TRB2_ARATH	Q39242 arabidopsis
411	7	0.5	386	1	RMAR_HANWI	P48849 hansenula w
412	7	0.5	387	1	WR36_ARATH	Q9car4 arabidopsis
413	7	0.5	389	1	SERI_BOMMO	P07856 bombyx mori
414	7	0.5	392	1	ODD_DROME	P23803 drosophila
415	7	0.5	393	1	IRTF_HUMAN	Q00978 homo sapien
416	7	0.5	394	1	VATC_SCHPO	Q9hdw6 schizosacch
417	7	0.5	397	1	HKL2_MALDO	O04135 malus domes
418	7	0.5	397	1	YKQ8_CAEEL	P34303 caenorhabdi
419	7	0.5	398	1	CAR4_RHINI	Q03700 rhizopus ni
420	7	0.5	399	1	HM39_CAEEL	Q22812 caenorhabdi
421	7	0.5	400	1	OMPA_THEMA	Q01969 thermotoga
422	7	0.5	401	1	NNOS_DROME	P25724 drosophila
423	7	0.5	406	1	SR40_YEAST	P32583 saccharomyc
424	7	0.5	407	1	ADRM_HUMAN	Q16186 homo sapien
425	7	0.5	407	1	ADRM_MOUSE	Q9jkv1 mus musculu
426	7	0.5	407	1	ADRM_RAT	Q9jmb5 rattus norv
427	7	0.5	407	1	RPI1_YEAST	P23250 saccharomyc
428	7	0.5	410	1	PGK_METFE	P20971 methanother
429	7	0.5	410	1	PO42_HUMAN	Q12837 homo sapien
430	7	0.5	410	1	VIE2_HCMVT	P06435 human cytom
431	7	0.5	411	1	PO42_MOUSE	Q63934 mus musculu
432	7	0.5	411	1	VIE2_HCMVA	P19893 human cytom
433	7	0.5	411	1	YA12_SCHPO	Q09685 schizosacch
434	7	0.5	414	1	NSR1_YEAST	P27476 saccharomyc
435	7	0.5	416	1	O35A_DROME	Q9v3q2 drosophila
436	7	0.5	416	1	SO_DROME	Q27350 drosophila
437	7	0.5	421	1	SYT1_MOUSE	P46096 mus musculu
438	7	0.5	421	1	SYT1_RAT	P21707 rattus norv
439	7	0.5	422	1	SYT1_BOVIN	P48018 bos taurus
440	7	0.5	422	1	SYT1_HUMAN	P21579 homo sapien
441	7	0.5	423	1	VDR_RAT	P13053 rattus norv
442	7	0.5	424	1	SYT1_CHICK	P47191 gallus gall
443	7	0.5	427	1	MYCN_SERCA	P26014 serinus can
444	7	0.5	427	1	SY61_DISOM	P24505 discopyge o
445	7	0.5	429	1	DR48_YEAST	P18899 saccharomyc
446	7	0.5	431	1	PO23_MOUSE	P31362 mus musculu
447	7	0.5	433	1	AFLR_EMENI	P52957 emericella
448	7	0.5	433	1	RTC1_DICDI	O15746 dictyosteli
449	7	0.5	434	1	V023_FOWPV	Q9j5h8 fowlpox vir
450	7	0.5	440	1	SOX4_MOUSE	Q06831 mus musculu
451	7	0.5	440	1	VRN2_ARATH	Q8w5b1 arabidopsis
452	7	0.5	444	1	FXF2_HUMAN	Q12947 homo sapien
453	7	0.5	449	1	TIG_UREPA	Q9pqe4 ureaplasma
454	7	0.5	450	1	THI2_YEAST	P38141 saccharomyc
455	7	0.5	450	1	UTH1_YEAST	P36135 saccharomyc
456	7	0.5	452	1	K1CO_MOUSE	Q61414 mus musculu
457	7	0.5	452	1	PUB1_YEAST	P32588 saccharomyc
458	7	0.5	453	1	MDHP_FLABI	P46489 flaveria bi
459	7	0.5	461	1	PLSB_PHAVU	Q43822 phaseolus v
460	7	0.5	462	1	SELA_CLOPE	Q8xik2 clostridium
461	7	0.5	464	1	CAP_DICDI	P54654 dictyosteli
462	7	0.5	464	1	IFE_BRALA	Q04948 branchiosto
463	7	0.5	465	1	HNF6_HUMAN	Q9ubc0 homo sapien
464	7	0.5	465	1	HNF6_MOUSE	O08755 mus musculu
465	7	0.5	465	1	HNF6_RAT	P70512 rattus norv
466	7	0.5	466	1	CYP8_CAEEL	P52016 caenorhabdi
467	7	0.5	468	1	KG3H_DICDI	P51136 dictyosteli
468	7	0.5	468	1	YL34_CAEEL	P34425 caenorhabdi
469	7	0.5	469	1	EC33_YEAST	P38248 saccharomyc
470	7	0.5	470	1	PLSB_CUCSA	Q39639 cucumis sat
471	7	0.5	471	1	HH_DROME	Q02936 drosophila
472	7	0.5	472	1	K1CP_HUMAN	P08779 homo sapien
473	7	0.5	472	1	PLSB_SPIOL	Q43869 spinacia ol
474	7	0.5	473	1	ATH1_ARATH	P48731 arabidopsis
475	7	0.5	474	1	LAM3_MOUSE	P48680 mus musculu
476	7	0.5	474	1	SOX4_HUMAN	Q06945 homo sapien
477	7	0.5	475	1	SIM1_YEAST	P40472 saccharomyc
478	7	0.5	476	1	CBP2_HORVU	P08818 hordeum vul
479	7	0.5	476	1	RF3_SACBA	P05512 saccharomyc
480	7	0.5	479	1	DBPA_BACSU	P42305 bacillus su
481	7	0.5	479	1	YP66_YEAST	Q12194 saccharomyc
482	7	0.5	481	1	HH_DROHY	P56674 drosophila
483	7	0.5	482	1	YSR2_CAEEL	Q09950 caenorhabdi
484	7	0.5	484	1	YOD0_YEAST	Q08193 saccharomyc
485	7	0.5	485	1	Y136_TREPA	O83172 treponema p
486	7	0.5	487	1	OAF_DROME	Q9nla6 drosophila
487	7	0.5	489	1	MDM2_MOUSE	P23804 mus musculu
488	7	0.5	489	1	SFR4_MOUSE	Q8ve97 mus musculu
489	7	0.5	489	1	WR47_ARATH	Q9zsi7 arabidopsis
490	7	0.5	491	1	KCS3_HUMAN	Q9bq31 homo sapien
491	7	0.5	491	1	KCS3_RABIT	Q9tt17 oryctolagus
492	7	0.5	491	1	KCS3_RAT	O88759 rattus norv
493	7	0.5	493	1	HMAB_DROME	P09087 drosophila
494	7	0.5	500	1	DCE_PETHY	Q07346 petunia hyb
495	7	0.5	501	1	ANKH_BRARE	P58368 brachydanio
496	7	0.5	503	1	WSC2_YEAST	P53832 saccharomyc
497	7	0.5	504	1	VL1_HPV51	P26536 human papil
498	7	0.5	505	1	GPMI_MYCPU	Q98qa7 mycoplasma
499	7	0.5	506	1	NPL3_HUMAN	Q99457 homo sapien
500	7	0.5	508	1	CROC_DROME	P32027 drosophila
501	7	0.5	510	1	KPK2_PLAFK	Q02595 plasmodium
502	7	0.5	510	1	MS2P_CRIGR	O54862 cricetulus
503	7	0.5	511	1	GUNB_PSEFL	P18126 pseudomonas
504	7	0.5	512	1	GLGT_VICFA	P52417 vicia faba
505	7	0.5	512	1	HMZ2_ARATH	O04921 arabidopsis
506	7	0.5	512	1	WR33_ARATH	Q8s8p5 arabidopsis
507	7	0.5	513	1	COX1_PAPHA	Q9zxy2 papio hamad
508	7	0.5	519	1	CBX2_MOUSE	P30658 mus musculu
509	7	0.5	519	1	MS2P_HUMAN	O43462 homo sapien
510	7	0.5	521	1	PTP1_DICDI	P34137 dictyosteli
511	7	0.5	524	1	HME2_SCHMA	Q26601 schistosoma
512	7	0.5	524	1	HUNB_TRICA	Q01791 tribolium c
513	7	0.5	528	1	YAC8_YEAST	P39734 saccharomyc
514	7	0.5	532	1	ICP0_HSVEB	P28990 equine herp
515	7	0.5	536	1	YEN1_SCHPO	O13695 schizosacch
516	7	0.5	537	1	SP70_DICDI	P15269 dictyosteli
517	7	0.5	538	1	IF3C_EUGGR	P36177 euglena gra
518	7	0.5	543	1	EGR1_HUMAN	P18146 homo sapien
519	7	0.5	543	1	HMA1_ARATH	P42804 arabidopsis
520	7	0.5	543	1	RRP3_YEAST	P38712 saccharomyc
521	7	0.5	543	1	THSB_THEAC	P48425 thermoplasm
522	7	0.5	546	1	PHR2_CANAL	O13318 candida alb
523	7	0.5	548	1	ERF_HUMAN	P50548 homo sapien
524	7	0.5	548	1	WR72_ARATH	Q9lxg8 arabidopsis
525	7	0.5	549	1	EPD2_CANMA	O74137 candida mal
526	7	0.5	551	1	ERF_MOUSE	P70459 mus musculu
527	7	0.5	551	1	YG1F_YEAST	P53214 saccharomyc
528	7	0.5	552	1	HMEN_DROME	P02836 drosophila
529	7	0.5	559	1	ENS2_YEAST	P12294 saccharomyc
530	7	0.5	559	1	GAS1_YEAST	P22146 saccharomyc
531	7	0.5	560	1	GAT2_YEAST	P40209 saccharomyc
532	7	0.5	560	1	P2B2_DROME	Q27889 drosophila
533	7	0.5	567	1	CHI3_CANAL	P40954 candida alb
534	7	0.5	568	1	AF9_HUMAN	P42568 homo sapien
535	7	0.5	569	1	ESR1_BRARE	P57717 brachydanio
536	7	0.5	571	1	XYNC_PSEFL	P23031 pseudomonas
537	7	0.5	573	1	YEY8_YEAST	P40095 saccharomyc
538	7	0.5	577	1	MTG8_MOUSE	Q61909 mus musculu
539	7	0.5	578	1	PRIM_BUCBP	Q89b09 buchnera ap
540	7	0.5	579	1	M3K7_MOUSE	Q62073 mus musculu
541	7	0.5	580	1	MDLB_BUCAP	Q8k984 buchnera ap
542	7	0.5	582	1	SYN2_HUMAN	Q92777 homo sapien
543	7	0.5	583	1	AXU1_MOUSE	P59054 mus musculu
544	7	0.5	583	1	CHI2_CANAL	P40953 candida alb

545	7	0.5	586	1	HM26_CAEEL	P34522	caenorhabdi
546	7	0.5	586	1	SYN2_RAT	Q63537	rattus norv
547	7	0.5	589	1	AXU1_HUMAN	Q96s65	homo sapien
548	7	0.5	590	1	HMAA_DROME	P29555	drosophila
549	7	0.5	590	1	NPA1_HUMAN	Q99742	homo sapien
550	7	0.5	592	1	ODP2_DICDI	P36413	dictyosteli
551	7	0.5	592	1	XYNB_PSEFL	P23030	pseudomonas
552	7	0.5	593	1	CC23_SCHPO	O42709	schizosacch
553	7	0.5	594	1	NPA1_MOUSE	P97459	mus musculu
554	7	0.5	594	1	RBJK_DROME	P28159	drosophila
555	7	0.5	598	1	RACA_DICDI	P34147	dictyosteli
556	7	0.5	600	1	LAM2_CHICK	P14732	gallus gall
557	7	0.5	600	1	XKS1_YEAST	P42826	saccharomyc
558	7	0.5	601	1	DEAD_BUCAI	P57453	buchnera ap
559	7	0.5	601	1	DEAD_BUCAP	Q8k9h6	buchnera ap
560	7	0.5	602	1	DEAD_BUCBP	Q89af9	buchnera ap
561	7	0.5	603	1	BUD8_YEAST	P41698	saccharomyc
562	7	0.5	604	1	DED1_YEAST	P06634	saccharomyc
563	7	0.5	604	1	MTG8_HUMAN	Q06455	homo sapien
564	7	0.5	605	1	WIS1_SCHPO	P33886	schizosacch
565	7	0.5	606	1	M3K7_HUMAN	O43318	homo sapien
566	7	0.5	610	1	SAN1_YEAST	P22470	saccharomyc
567	7	0.5	611	1	XYNA_PSEFL	P14768	pseudomonas
568	7	0.5	613	1	DEAD_HAEIN	P44586	haemophilus
569	7	0.5	614	1	DDX5_HUMAN	P17844	homo sapien
570	7	0.5	614	1	DDX5_MOUSE	Q61656	mus musculu
571	7	0.5	614	1	NRD1_HUMAN	P20393	homo sapien
572	7	0.5	617	1	DBP1_YEAST	P24784	saccharomyc
573	7	0.5	618	1	ORC2_DROME	Q24168	drosophila
574	7	0.5	618	1	ZEST_DROVI	Q24762	drosophila
575	7	0.5	619	1	DNAK_CLOPE	P26823	clostridium
576	7	0.5	621	1	PPID_BUCAP	Q8k987	buchnera ap
577	7	0.5	622	1	YRT1_CAEEL	Q10044	caenorhabdi
578	7	0.5	624	1	PBPD_BACSU	P40750	bacillus su
579	7	0.5	624	1	RN12_HUMAN	Q9nvw2	homo sapien
580	7	0.5	628	1	DEAD_ECOLI	P23304	escherichia
581	7	0.5	632	1	ARPB_ECOLI	P76205	escherichia
582	7	0.5	633	1	CNE7_HUMAN	Q9ub16	homo sapien
583	7	0.5	634	1	DNK2_SYNP7	P50021	synechococc
584	7	0.5	634	1	HWP1_CANAL	P46593	candida alb
585	7	0.5	634	1	Z297_HUMAN	O15209	homo sapien
586	7	0.5	636	1	DED1_SCHPO	O13370	schizosacch
587	7	0.5	638	1	K22O_HUMAN	Q01546	homo sapien
588	7	0.5	640	1	Y551_SYNY3	P54123	synechocyst
589	7	0.5	641	1	FIB2_PETMA	P33573	petromyzon
590	7	0.5	642	1	DEAD_KLEPN	P33906	klebsiella
591	7	0.5	644	1	BTD_DROME	Q24266	drosophila
592	7	0.5	649	1	LONH_METJA	Q58812	methanococc
593	7	0.5	650	1	DD17_HUMAN	Q92841	homo sapien
594	7	0.5	650	1	Y411_HUMAN	O43295	homo sapien
595	7	0.5	652	1	DREB_CHICK	P18302	gallus gall
596	7	0.5	654	1	HS77_YEAST	P12398	saccharomyc
597	7	0.5	656	1	ACSA_RHOCA	O68040	rhodobacter
598	7	0.5	656	1	YNE7_YEAST	P53955	saccharomyc
599	7	0.5	657	1	GLGX_ECOLI	P15067	escherichia
600	7	0.5	660	1	DDXY_HUMAN	O15523	homo sapien
601	7	0.5	660	1	PL10_MOUSE	P16381	mus musculu
602	7	0.5	661	1	DDX3_HUMAN	O00571	homo sapien
603	7	0.5	661	1	DDX3_MOUSE	Q62167	mus musculu
604	7	0.5	661	1	VASA_DROME	P09052	drosophila
605	7	0.5	662	1	MBB1_CHLRE	Q9fns4	chlamydomon
606	7	0.5	662	1	YFL8_YEAST	P43620	saccharomyc
607	7	0.5	663	1	PIT_DROME	Q9vd51	drosophila
608	7	0.5	663	1	YJO9_YEAST	P47005	saccharomyc
609	7	0.5	665	1	KF22_HUMAN	Q14807	homo sapien
610	7	0.5	666	1	NOD_DROME	P18105	drosophila
611	7	0.5	671	1	CHS5_YEAST	Q12114	saccharomyc
612	7	0.5	677	1	HSF_KLULA	P22121	kluyveromyc
613	7	0.5	677	1	UBF1_XENLA	P25979	xenopus lae
614	7	0.5	679	1	PAN3_YEAST	P36102	saccharomyc
615	7	0.5	684	1	EP84_HCMVA	P17151	human cytom
616	7	0.5	684	1	RPF1_HUMAN	P78424	homo sapien
617	7	0.5	686	1	PLB3_YEAST	Q08108	saccharomyc
618	7	0.5	689	1	KF1B_RAT	O88658	rattus norv
619	7	0.5	690	1	GSH1_CANAL	Q9hf78	candida alb
620	7	0.5	693	1	CAUP_DROME	P54269	drosophila
621	7	0.5	697	1	AN3_XENLA	P24346	xenopus lae
622	7	0.5	699	1	KI22_STRPU	P46872	strongyloce
623	7	0.5	699	1	NP14_HUMAN	Q14978	homo sapien
624	7	0.5	699	1	SRCH_HUMAN	P23327	homo sapien
625	7	0.5	701	1	KF3A_MOUSE	P28741	mus musculu
626	7	0.5	701	1	UBF2_XENLA	P25980	xenopus lae
627	7	0.5	702	1	DDX4_MOUSE	Q61496	mus musculu
628	7	0.5	702	1	KF3A_HUMAN	Q9y496	homo sapien
629	7	0.5	704	1	CT67_HUMAN	Q9h4z3	homo sapien
630	7	0.5	704	1	NP14_RAT	P41777	rattus norv
631	7	0.5	706	1	CIAA_PAEPP	Q45358	paenibacill
632	7	0.5	706	1	CT67_MOUSE	P59114	mus musculu
633	7	0.5	706	1	DREB_MOUSE	Q9qxs6	mus musculu
634	7	0.5	706	1	FTSH_MEDSA	Q9bae0	medicago sa
635	7	0.5	706	1	KIP2_YEAST	P28743	saccharomyc
636	7	0.5	707	1	DREB_RAT	Q07266	rattus norv
637	7	0.5	709	1	PPZ2_YEAST	P33329	saccharomyc
638	7	0.5	710	1	GSQD_ERWCH	Q01565	erwinia chr
639	7	0.5	712	1	GSPD_ERWCH	P31700	erwinia chr
640	7	0.5	713	1	DDX4_RAT	Q64060	rattus norv
641	7	0.5	715	1	ADSV_BOVIN	Q28046	bos taurus
642	7	0.5	720	1	GLH3_CAEEL	O01836	caenorhabdi
643	7	0.5	720	1	NRG3_HUMAN	P56975	homo sapien
644	7	0.5	723	1	MASZ_RHILV	Q937w7	rhizobium l
645	7	0.5	723	1	MASZ_RHIME	Q92ta4	rhizobium m
646	7	0.5	724	1	DDX4_HUMAN	Q9nqi0	homo sapien
647	7	0.5	727	1	MFP1_ARATH	Q9lw85	arabidopsis
648	7	0.5	727	1	PKP1_BOVIN	Q28161	bos taurus
649	7	0.5	728	1	VIV_ORYSA	P37398	oryza sativ
650	7	0.5	730	1	CCT2_HUMAN	O60583	homo sapien
651	7	0.5	730	1	CDK9_CAEEL	P46551	caenorhabdi
652	7	0.5	730	1	GLN3_YEAST	P18494	saccharomyc
653	7	0.5	731	1	BAF1_YEAST	P14164	saccharomyc
654	7	0.5	733	1	RRP6_YEAST	Q12149	saccharomyc
655	7	0.5	741	1	NFL1_MOUSE	Q61985	mus musculu
656	7	0.5	742	1	DBP7_YEAST	P36120	saccharomyc
657	7	0.5	742	1	KI21_STRPU	P46871	strongyloce
658	7	0.5	742	1	KM65_YEAST	Q03656	saccharomyc
659	7	0.5	746	1	GYP7_YEAST	P48365	saccharomyc
660	7	0.5	747	1	KF3B_HUMAN	O15066	homo sapien
661	7	0.5	747	1	KF3B_MOUSE	Q61771	mus musculu
662	7	0.5	747	1	PARC_HAEIN	P43702	haemophilus
663	7	0.5	748	1	GUNC_PSEFL	P27033	pseudomonas
664	7	0.5	748	1	KHL1_HUMAN	Q9nr64	homo sapien
665	7	0.5	751	1	KHL1_MOUSE	Q9ji74	mus musculu
666	7	0.5	753	1	TKR_DROME	P14083	drosophila
667	7	0.5	754	1	ASPH_BOVIN	Q28056	bos taurus
668	7	0.5	757	1	2A5D_YEAST	P38903	saccharomyc
669	7	0.5	760	1	AMY_CLOAB	P23671	clostridium
670	7	0.5	763	1	GLH1_CAEEL	P34689	caenorhabdi
671	7	0.5	764	1	CSW_DROVI	Q24708	drosophila
672	7	0.5	764	1	UBF1_HUMAN	P17480	homo sapien
673	7	0.5	764	1	UBF1_RAT	P25977	rattus norv
674	7	0.5	765	1	UBF1_MOUSE	P25976	mus musculu
675	7	0.5	772	1	NFL1_HUMAN	Q14494	h nuclear f
676	7	0.5	776	1	KLP1_CHLRE	P46870	chlamydomon
677	7	0.5	779	1	NIM1_NEUCR	P48479	neurospora
678	7	0.5	783	1	KMHC_DICDI	P34125	dictyosteli
679	7	0.5	785	1	SOK2_YEAST	P53438	saccharomyc
680	7	0.5	786	1	FL10_CHLRE	P46869	chlamydomon
681	7	0.5	789	1	AD07_RAT	Q63180	rattus norv
682	7	0.5	790	1	KIF9_HUMAN	Q9haq2	homo sapien
683	7	0.5	790	1	KIF9_MOUSE	Q9wv04	mus musculu
684	7	0.5	791	1	KF17_HUMAN	Q9p2e2	homo sapien
685	7	0.5	793	1	KF3C_HUMAN	O14782	homo sapien
686	7	0.5	795	1	TBC5_HUMAN	Q92609	homo sapien
687	7	0.5	796	1	KF3C_MOUSE	O35066	mus musculu
688	7	0.5	796	1	KF3C_RAT	O55165	rattus norv
689	7	0.5	801	1	BRD2_HUMAN	P25440	homo sapien
690	7	0.5	805	1	DF19_CAEEL	Q09555	caenorhabdi

691	7	0.5	805	1	E2F_DROME	Q27368 drosophila
692	7	0.5	805	1	PIF1_SCHPO	Q9uua2 schizosacch
693	7	0.5	807	1	YAK1_YEAST	P14680 saccharomyc
694	7	0.5	808	1	FTFB_DROME	Q05192 drosophila
695	7	0.5	810	1	REB1_YEAST	P21538 saccharomyc
696	7	0.5	816	1	HUNB_DROVI	P13361 drosophila
697	7	0.5	816	1	YG3A_YEAST	P53278 saccharomyc
698	7	0.5	817	1	VRP1_YEAST	P37370 saccharomyc
699	7	0.5	829	1	E74A_DROME	P20105 drosophila
700	7	0.5	838	1	L100_ADEP3	Q9ytr7 porcine ade
701	7	0.5	843	1	CO7_PIG	Q9tuq3 sus scrofa
702	7	0.5	843	1	CYP1_BRUMA	Q27450 brugia mala
703	7	0.5	845	1	CLPC_CHLPN	Q9z8a6 chlamydia p
704	7	0.5	845	1	ITBX_DROME	P11584 drosophila
705	7	0.5	850	1	D7_DICDI	P54682 dictyosteli
706	7	0.5	850	1	NRG2_HUMAN	O14511 homo sapien
707	7	0.5	854	1	CC24_YEAST	P11433 saccharomyc
708	7	0.5	856	1	KF23_HUMAN	Q02241 homo sapien
709	7	0.5	859	1	ALR1_YEAST	Q08269 saccharomyc
710	7	0.5	864	1	YC18_HUMAN	Q9ulk2 homo sapien
711	7	0.5	866	1	MYSP_SCHMA	P06198 schistosoma
712	7	0.5	868	1	VAV2_MOUSE	Q60992 mus musculu
713	7	0.5	878	1	KPCO_HUMAN	Q9bzl6 homo sapien
714	7	0.5	878	1	VAV2_HUMAN	P52735 homo sapien
715	7	0.5	880	1	BRC4_DROME	Q24206 drosophila
716	7	0.5	881	1	PRY3_YEAST	P47033 saccharomyc
717	7	0.5	883	1	E74B_DROME	P11536 drosophila
718	7	0.5	885	1	YD03_YEAST	Q06639 saccharomyc
719	7	0.5	887	1	K20A_MOUSE	P97329 mus musculu
720	7	0.5	890	1	K20A_HUMAN	O95235 homo sapien
721	7	0.5	892	1	ATX7_HUMAN	O15265 homo sapien
722	7	0.5	894	1	WPRA_BACSU	P54423 bacillus su
723	7	0.5	902	1	RNE_BUCAI	P57429 buchnera ap
724	7	0.5	903	1	DPOL_BPR69	Q38087 bacteriopha
725	7	0.5	909	1	MTC2_YEAST	P53185 saccharomyc
726	7	0.5	918	1	PMA1_YEAST	P05030 saccharomyc
727	7	0.5	925	1	PDC2_YEAST	P32896 saccharomyc
728	7	0.5	928	1	KINH_NEUCR	P48467 neurospora
729	7	0.5	928	1	KKR1_YEAST	P36003 saccharomyc
730	7	0.5	936	1	FHL1_YEAST	P39521 saccharomyc
731	7	0.5	939	1	ST20_YEAST	Q03497 saccharomyc
732	7	0.5	939	1	XPC_HUMAN	Q01831 homo sapien
733	7	0.5	946	1	YBT6_YEAST	P38250 saccharomyc
734	7	0.5	947	1	PMA2_YEAST	P19657 saccharomyc
735	7	0.5	950	1	URB1_USTMA	P40349 ustilago ma
736	7	0.5	958	1	MSH4_MOUSE	Q99mt2 mus musculu
737	7	0.5	966	1	FIB1_PETMA	P02674 petromyzon
738	7	0.5	971	1	CLA4_CANAL	O14427 candida alb
739	7	0.5	972	1	TOP1_DROME	P30189 drosophila
740	7	0.5	980	1	KFMS_FELCA	P13369 felis silve
741	7	0.5	982	1	CRM_DROME	O76906 drosophila
742	7	0.5	989	1	SERA_PLAFG	P13823 plasmodium
743	7	0.5	993	1	RPOT_ARATH	O24600 arabidopsis
744	7	0.5	1004	1	SAL2_MOUSE	Q9qx96 mus musculu
745	7	0.5	1006	1	EPB6_HUMAN	O15197 homo sapien
746	7	0.5	1006	1	K125_TOBAC	O23826 nicotiana t
747	7	0.5	1007	1	SAL2_HUMAN	Q9y467 homo sapien
748	7	0.5	1010	1	WNT5_DROME	P28466 drosophila
749	7	0.5	1025	1	KG09_HUMAN	Q9h0a0 homo sapien
750	7	0.5	1029	1	KSP1_YEAST	P38691 saccharomyc
751	7	0.5	1029	1	RIP3_RAT	Q9ere6 rattus norv
752	7	0.5	1036	1	NIT2_NEUCR	P19212 neurospora
753	7	0.5	1047	1	HIRA_DROME	O17468 drosophila
754	7	0.5	1050	1	ULK1_HUMAN	O75385 homo sapien
755	7	0.5	1051	1	ULK1_MOUSE	O70405 mus musculu
756	7	0.5	1052	1	MGPC_MYCGE	P22747 mycoplasma
757	7	0.5	1056	1	K125_ARATH	P82266 arabidopsis
758	7	0.5	1057	1	GCSP_PEA	P26969 pisum sativ
759	7	0.5	1057	1	KF11_HUMAN	P52732 homo sapien
760	7	0.5	1060	1	EG51_XENLA	P28025 xenopus lae
761	7	0.5	1065	1	SED4_YEAST	P25365 saccharomyc
762	7	0.5	1066	1	KL61_DROME	P46863 drosophila
763	7	0.5	1067	1	EG52_XENLA	Q91783 xenopus lae
764	7	0.5	1070	1	AGLU_CANTS	P29064 candida tsu
765	7	0.5	1075	1	FLO5_YEAST	P38894 saccharomyc
766	7	0.5	1077	1	HLES_DROME	Q02308 drosophila
767	7	0.5	1081	1	GALY_YEAST	P19659 saccharomyc
768	7	0.5	1082	1	SP23_YEAST	P35210 saccharomyc
769	7	0.5	1085	1	AMP1_PLAFQ	O96935 plasmodium
770	7	0.5	1085	1	CUT7_SCHPO	P24339 schizosacch
771	7	0.5	1090	1	NIT4_NEUCR	P28349 neurospora
772	7	0.5	1093	1	AF17_HUMAN	P55198 homo sapien
773	7	0.5	1097	1	CCT_DROME	O96433 drosophila
774	7	0.5	1097	1	KF1D_RAT	O35787 rattus norv
775	7	0.5	1103	1	KF1C_HUMAN	O43896 homo sapien
776	7	0.5	1111	1	KIP1_YEAST	P28742 saccharomyc
777	7	0.5	1115	1	TBC2_CHLRE	Q8vxp3 chlamydomon
778	7	0.5	1141	1	SRE2_HUMAN	Q12772 homo sapien
779	7	0.5	1142	1	MGC1_HUMAN	O60732 homo sapien
780	7	0.5	1142	1	PAK1_YEAST	P38990 saccharomyc
781	7	0.5	1146	1	YHC3_YEAST	P38742 saccharomyc
782	7	0.5	1147	1	KIN2_YEAST	P13186 saccharomyc
783	7	0.5	1156	1	PHYB_SOYBN	P42499 glycine max
784	7	0.5	1169	1	YK82_YEAST	P36170 saccharomyc
785	7	0.5	1184	1	BIMC_EMENI	P17120 emericella
786	7	0.5	1185	1	DRPL_HUMAN	P54259 homo sapien
787	7	0.5	1189	1	YJH6_YEAST	P47035 saccharomyc
788	7	0.5	1202	1	RPM2_YEAST	Q02773 saccharomyc
789	7	0.5	1210	1	AF4_HUMAN	P51825 homo sapien
790	7	0.5	1211	1	BUN2_DROME	Q24523 drosophila
791	7	0.5	1214	1	BRF1_HUMAN	P55201 homo sapien
792	7	0.5	1224	1	ABL1_CAEEL	P03949 caenorhabdi
793	7	0.5	1227	1	LAF4_HUMAN	P51826 homo sapien
794	7	0.5	1235	1	TRK1_YEAST	P12685 saccharomyc
795	7	0.5	1237	1	E75A_DROME	P17671 drosophila
796	7	0.5	1238	1	YN13_YEAST	P53840 saccharomyc
797	7	0.5	1244	1	SLA1_YEAST	P32790 saccharomyc
798	7	0.5	1271	1	Y338_MYCGE	P47580 mycoplasma
799	7	0.5	1298	1	ICP4_HSV11	P08392 herpes simp
800	7	0.5	1310	1	ICP4_VZVD	P09310 varicella-z
801	7	0.5	1324	1	SAL1_HUMAN	Q9nsc2 homo sapien
802	7	0.5	1328	1	HUS2_SCHPO	Q09811 schizosacch
803	7	0.5	1337	1	YDM5_SCHPO	P87136 schizosacch
804	7	0.5	1341	1	ACIN_HUMAN	Q9ukv3 homo sapien
805	7	0.5	1394	1	E75B_DROME	P17672 drosophila
806	7	0.5	1396	1	ITA2_DROME	P12080 drosophila
807	7	0.5	1402	1	SALM_DROVI	P39806 drosophila
808	7	0.5	1402	1	Y232_HUMAN	Q92628 homo sapien
809	7	0.5	1415	1	ICP4_HSVMG	Q02362 marek's dis
810	7	0.5	1443	1	E75C_DROME	P13055 drosophila
811	7	0.5	1444	1	ADP1_MYCGE	P20796 mycoplasma
812	7	0.5	1446	1	IE18_PRVKA	P33479 pseudorabie
813	7	0.5	1461	1	IE18_PRVIF	P11675 pseudorabie
814	7	0.5	1487	1	ICP4_HSVEB	P28925 equine herp
815	7	0.5	1487	1	ICP4_HSVEK	P17473 equine herp
816	7	0.5	1489	1	YGP0_YEAST	P53115 saccharomyc
817	7	0.5	1509	1	GSR1_HUMAN	Q9nzm4 homo sapien
818	7	0.5	1531	1	YQ38_CAEEL	Q09459 caenorhabdi
819	7	0.5	1537	1	DNM1_CHICK	Q92072 gallus gall
820	7	0.5	1579	1	SSK2_YEAST	P53599 saccharomyc
821	7	0.5	1581	1	PPRB_HUMAN	Q15648 h peroxisom
822	7	0.5	1603	1	PSC_DROME	P35820 drosophila
823	7	0.5	1609	1	FIG2_YEAST	P25653 saccharomyc
824	7	0.5	1627	1	ADP1_MYCPN	P11311 mycoplasma
825	7	0.5	1659	1	VIT_ONCMY	Q92093 oncorhynchu
826	7	0.5	1687	1	VIT2_FUNHE	Q98893 fundulus he
827	7	0.5	1690	1	KF1A_HUMAN	Q12756 homo sapien
828	7	0.5	1695	1	KF1A_MOUSE	P33173 mus musculu
829	7	0.5	1719	1	PRD2_HUMAN	Q13029 homo sapien
830	7	0.5	1723	1	PM20_CHLPN	Q9z812 chlamydia p
831	7	0.5	1772	1	MSP1_PLAYO	P13828 plasmodium
832	7	0.5	1782	1	VIT_BOMMO	Q27309 bombyx mori
833	7	0.5	1807	1	VTA2_XENLA	P18709 xenopus lae
834	7	0.5	1816	1	KF1B_HUMAN	O60333 homo sapien
835	7	0.5	1816	1	KF1B_MOUSE	Q60575 mus musculu
836	7	0.5	1822	1	ZAP3_HUMAN	P49750 homo sapien

837	7	0.5	1823	1	VIT_ICHUN	Q91062	ichthyomyzo
838	7	0.5	1882	1	Y468_MYCPN	P75109	mycoplasma
839	7	0.5	1912	1	VIT1_CHICK	P87498	gallus gall
840	7	0.5	1956	1	ATX1_PLAFA	Q04956	plasmodium
841	7	0.5	1960	1	TF20_HUMAN	Q9ugu0	homo sapien
842	7	0.5	1983	1	TF20_MOUSE	Q9epq8	mus musculu
843	7	0.5	2004	1	CHDB_HUMAN	Q9hck8	homo sapien
844	7	0.5	2038	1	FSH_DROME	P13709	drosophila
845	7	0.5	2148	1	VIT1_AEDAE	Q16927	aedes aegyp
846	7	0.5	2167	1	BEM2_YEAST	P39960	saccharomyc
847	7	0.5	2174	1	T240_HUMAN	Q9uhv7	homo sapien
848	7	0.5	2182	1	CABI_RAT	O88480	rattus norv
849	7	0.5	2220	1	CABI_HUMAN	Q9y6j0	homo sapien
850	7	0.5	2505	1	CCAA_HUMAN	O00555	homo sapien
851	7	0.5	2663	1	CENE_HUMAN	Q02224	homo sapien
852	7	0.5	3097	1	CADN_DROME	O15943	drosophila
853	7	0.5	3174	1	CHAC_HUMAN	Q96r17	homo sapien
854	7	0.5	3255	1	POLG_LMV0	P31999	l genome po
855	7	0.5	3255	1	POLG_LMVE	P89876	l genome po
856	7	0.5	3969	1	HRX_HUMAN	Q03164	homo sapien
857	7	0.5	4705	1	FAT2_DROME	Q9vw71	drosophila
858	7	0.5	5147	1	FAT_DROME	P33450	drosophila
859	6	0.5	37	1	RK36_EUGGR	P21532	euglena gra
860	6	0.5	40	1	VIT_MELGA	P56531	meleagris g
861	6	0.5	50	1	R331_MYCPU	Q98q77	mycoplasma
862	6	0.5	65	1	V07K_CLV	P28898	carnation l
863	6	0.5	67	1	YQ12_BACAN	Q9rn20	bacillus an
864	6	0.5	69	1	RL38_HUMAN	P23411	homo sapien
865	6	0.5	69	1	RL38_LYCES	P46291	lycopersico
866	6	0.5	69	1	RL38_MOUSE	Q9jji8	mus musculu
867	6	0.5	73	1	RL7_STAAU	P48860	staphylococ
868	6	0.5	74	1	CHI1_STROI	P29115	streptomyce
869	6	0.5	77	1	S18M_HUMAN	Q9uha2	homo sapien
870	6	0.5	78	1	KCRU_PIG	Q29577	sus scrofa
871	6	0.5	79	1	RS18_UREPA	Q9ppt8	ureaplasma
872	6	0.5	80	1	DBHL_RICCN	Q92h14	rickettsia
873	6	0.5	80	1	DBHL_RICPR	Q9zd26	rickettsia
874	6	0.5	80	1	IF1_DEIRA	Q9rsk1	deinococcus
875	6	0.5	80	1	YSPT_SHIFL	Q55298	shigella fl
876	6	0.5	88	1	YE4D_SCHPO	O13973	schizosacch
877	6	0.5	89	1	SLTB_BPH30	P08027	bacteriopha
878	6	0.5	89	1	Y150_ARCFU	O30087	archaeoglob
879	6	0.5	89	1	YE96_CLOPE	Q8xka4	clostridium
880	6	0.5	90	1	DBHL_RICRI	O68451	rickettsia
881	6	0.5	90	1	HSB7_RAT	Q9quk5	rattus norv
882	6	0.5	90	1	Y060_NPVOP	O10317	orgyia pseu
883	6	0.5	91	1	SMT3_CAEEL	P55853	caenorhabdi
884	6	0.5	97	1	IM8A_HUMAN	O60220	homo sapien
885	6	0.5	97	1	IM8A_RAT	Q9wva1	rattus norv
886	6	0.5	103	1	KAC5_RABIT	P01841	oryctolagus
887	6	0.5	104	1	YJB8_YEAST	P47070	saccharomyc
888	6	0.5	105	1	V249_FOWPV	P14363	fowlpox vir
889	6	0.5	105	1	Y613_ARCFU	O29642	archaeoglob
890	6	0.5	106	1	RNT1_TRIHA	P26875	trichoderma
891	6	0.5	106	1	THIO_GEOCY	O96952	geodia cydo
892	6	0.5	107	1	FKB1_XENLA	O42123	xenopus lae
893	6	0.5	107	1	YP94_YEAST	Q06835	saccharomyc
894	6	0.5	108	1	RL23_MYCGA	O52334	mycoplasma
895	6	0.5	108	1	SVS4_MOUSE	P18419	mus musculu
896	6	0.5	109	1	PRVA_AMPME	P02626	amphiuma me
897	6	0.5	111	1	RNPA_MYCPE	Q8eu90	mycoplasma
898	6	0.5	111	1	YPJD_BACSU	P42979	bacillus su
899	6	0.5	113	1	RL30_SPOFR	P58375	spodoptera
900	6	0.5	114	1	PHS_SULSO	Q97wm6	sulfolobus
901	6	0.5	114	1	YOJ5_CAEEL	P34628	caenorhabdi
902	6	0.5	115	1	A62F_DROME	O46202	drosophila
903	6	0.5	115	1	KV02_RABIT	P01683	oryctolagus
904	6	0.5	116	1	TCLA_MOUSE	P56280	mus musculu
905	6	0.5	118	1	YE17_SYNY3	P72731	synechocyst
906	6	0.5	118	1	YEJL_SCHPO	Q92363	schizosacch
907	6	0.5	121	1	H2B1_TETTH	P08993	tetrahymena
908	6	0.5	121	1	H2B2_TETTH	P08994	tetrahymena
909	6	0.5	121	1	OMP7_STAAU	P21223	staphylococ
910	6	0.5	122	1	MPP2_PHLPR	P43214	phleum prat
911	6	0.5	122	1	PA2_VIPBB	P31854	vipera beru
912	6	0.5	122	1	RL7_STAAM	Q99w66	staphylococ
913	6	0.5	122	1	YL15_ADE41	P23692	human adeno
914	6	0.5	124	1	CC26_YEAST	P14724	saccharomyc
915	6	0.5	124	1	RNP_CAMDR	P00670	camelus dro
916	6	0.5	125	1	Y394_RICPR	Q9zdd7	rickettsia
917	6	0.5	125	1	YE75_METJA	Q58870	methanococc
918	6	0.5	126	1	YBGS_ECOLI	P75758	escherichia
919	6	0.5	128	1	Y050_BPT4	P13303	bacteriopha
920	6	0.5	130	1	OAZ_DROVI	O44535	drosophila
921	6	0.5	130	1	RS8_BUCAP	P59030	buchnera ap
922	6	0.5	131	1	V092_FOWPV	Q9j5c8	fowlpox vir
923	6	0.5	131	1	YQHP_BACSU	P54514	bacillus su
924	6	0.5	132	1	RS19_PYRAB	Q9v1t9	pyrococcus
925	6	0.5	132	1	RS19_PYRFU	Q8u002	pyrococcus
926	6	0.5	132	1	RS19_PYRHO	O59422	pyrococcus
927	6	0.5	133	1	RBFA_CHLMU	Q9pku1	chlamydia m
928	6	0.5	133	1	YIM5_BPPH1	P10429	bacteriopha
929	6	0.5	134	1	RS24_YEAST	P26782	saccharomyc
930	6	0.5	135	1	YNP0_YEAST	P53902	saccharomyc
931	6	0.5	137	1	UCR7_SCHPO	O74533	schizosacch
932	6	0.5	137	1	YEDD_ECOLI	P31063	escherichia
933	6	0.5	138	1	ATPE_WIGBR	Q8d3j2	wiggleswort
934	6	0.5	138	1	HEX9_ADE07	P03283	human adeno
935	6	0.5	138	1	IF1A_SCHPO	P55877	schizosacch
936	6	0.5	139	1	YN06_YEAST	P53842	saccharomyc
937	6	0.5	139	1	YVYF_BACSU	P39807	bacillus su
938	6	0.5	140	1	R17D_ARATH	Q91z17	arabidopsis
939	6	0.5	140	1	RR8_EUGGR	P21508	euglena gra
940	6	0.5	140	1	YF62_ARCFU	O28710	archaeoglob
941	6	0.5	141	1	PER_DROSR	Q04537	drosophila
942	6	0.5	141	1	RNBR_GIRCA	Q29542	giraffa cam
943	6	0.5	142	1	DIM1_HUMAN	O14834	homo sapien
944	6	0.5	142	1	DIM1_SCHPO	P87215	schizosacch
945	6	0.5	142	1	Y742_CHLPN	Q9z7g3	chlamydia p
946	6	0.5	145	1	HA17_CLOBO	P46083	clostridium
947	6	0.5	145	1	YE1A_METJA	P81328	methanococc
948	6	0.5	145	1	YPH2_MYCCA	P45614	mycoplasma
949	6	0.5	146	1	ADF6_ARATH	Q9zsk2	arabidopsis
950	6	0.5	146	1	VG4_SPV1R	P15895	spiroplasma
951	6	0.5	147	1	SPAM_SALTY	P40612	salmonella
952	6	0.5	147	1	Y211_MYCGE	P47453	mycoplasma
953	6	0.5	147	1	YD05_METJA	Q58701	methanococc
954	6	0.5	148	1	GM6_TRYBG	Q26755	trypanosoma
955	6	0.5	148	1	PTR1_METJA	Q57615	methanococc
956	6	0.5	148	1	YG4E_YEAST	P42937	saccharomyc
957	6	0.5	149	1	BTT1_YEAST	P40314	saccharomyc
958	6	0.5	149	1	YM66_CLOPE	Q8xi55	clostridium
959	6	0.5	150	1	YHY5_SCHPO	O60154	schizosacch
960	6	0.5	152	1	FLIJ_BUCAP	Q8ka41	buchnera ap
961	6	0.5	152	1	Y447_WIGBR	Q8d2a7	wiggleswort
962	6	0.5	153	1	VE6_HPV2A	P25484	human papil
963	6	0.5	154	1	FASC_ECOLI	P45999	escherichia
964	6	0.5	154	1	PFD5_SCHPO	O94307	schizosacch
965	6	0.5	155	1	ITRF_MAIZE	P01088	zea mays (m
966	6	0.5	155	1	PPI1_SCHPO	P87051	schizosacch
967	6	0.5	156	1	RNP_MYOGL	Q9wus1	myoxus glis
968	6	0.5	156	1	YB01_RHOPA	O07465	rhodopseudo
969	6	0.5	157	1	PA1F_HUMAN	P24666	homo sapien
970	6	0.5	157	1	PA1S_HUMAN	P24667	homo sapien
971	6	0.5	157	1	PPAC_BOVIN	P11064	bos taurus
972	6	0.5	157	1	PPAC_PIG	P81693	sus scrofa
973	6	0.5	157	1	YWM3_CAEEL	Q23679	caenorhabdi
974	6	0.5	159	1	PPAC_RAT	P41498	rattus norv
975	6	0.5	161	1	YP49_RALSO	Q8xwc5	ralstonia s
976	6	0.5	161	1	YU12_CLOPE	Q93md5	clostridium
977	6	0.5	162	1	RL30_AERPE	Q9yf96	aeropyrum p
978	6	0.5	162	1	Y79A_METJA	P81313	methanococc
979	6	0.5	163	1	YA53_METJA	Q58453	methanococc
980	6	0.5	164	1	HIT1_YEAST	P46973	saccharomyc
981	6	0.5	167	1	HRS1_MOUSE	Q9qzu4	mus musculu
982	6	0.5	167	1	ISPF_BUCAP	Q8k9d7	buchnera ap

```
  983       6    0.5    167  1  PSAG_SPIOL                P12357 spinacia ol
  984       6    0.5    167  1  RNBR_BOVIN                P39873 bos taurus
  985       6    0.5    168  1  YNP7_YEAST                P53897 saccharomyc
  986       6    0.5    169  1  HSB7_MOUSE                P35385 mus musculu
  987       6    0.5    169  1  NUGC_MARPO                P12199 marchantia
  988       6    0.5    169  1  YGGD_ECOLI                P11663 escherichia
  989       6    0.5    170  1  RM06_DICDI                O21037 dictyosteli
  990       6    0.5    171  1  RS4_SULAC                 P39467 sulfolobus
  991       6    0.5    171  1  TRPP_CLOAB                Q97d63 clostridium
  992       6    0.5    171  1  Y328_AQUAE                O66665 aquifex aeo
  993       6    0.5    171  1  YTUX_AGRVI                P70790 agrobacteri
  994       6    0.5    172  1  INT2_BOVIN                P56830 bos taurus
  995       6    0.5    172  1  INT3_BOVIN                P56831 bos taurus
  996       6    0.5    173  1  SP12_YEAST                P17123 saccharomyc
  997       6    0.5    173  1  THM3_ARATH                Q9seu7 arabidopsis
  998       6    0.5    173  1  YNN3_YEAST                P53913 saccharomyc
  999       6    0.5    174  1  LEPH_STAAM                P72364 staphylococ
 1000       6    0.5    174  1  NU6M_ANOGA                P34856 anopheles g
```

ALIGNMENTS

```
RESULT 1
YEA7_YEAST
ID   YEA7_YEAST     STANDARD;      PRT;   666 AA.
AC   P40002;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 72.5 kDa protein in GCN4-WBP1 intergenic region.
GN   YEL007W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169868;
RA   Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA   Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA   Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA   Hunicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA   Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA   Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA   Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL   Nature 387:78-81(1997).
CC   -!- SIMILARITY: STRONG, TO YEAST YHR177W.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U18530; AAB64484.1; -.
DR   PIR; S50452; S50452.
DR   SGD; S0000733; TOS9.
KW   Hypothetical protein.
FT   DOMAIN      334    382       ASN-RICH.
SQ   SEQUENCE   666 AA;  72533 MW;  DC36F887BF2D17E0 CRC64;

  Query Match            1.3%;  Score 17;  DB 1;  Length 666;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-08;
  Matches   17;  Conservative   0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        351 NNNNNNNNNNNNNNNNI 367
```

```
RESULT 2
CC27_YEAST
ID   CC27_YEAST     STANDARD;      PRT;   758 AA.
AC   P38042;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cell division control protein 27.
GN   CDC27 OR SNB1 OR YBL084C OR YBL0718.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=96076635; PubMed=7502586;
RA   Obermaier B., Gassenhuber J., Piravandi E., Domdey H.;
RT   "Sequence analysis of a 78.6 kb segment of the left end of
RT   Saccharomyces cerevisiae chromosome II.";
RL   Yeast 11:1103-1112(1995).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A., AND MUTAGENESIS.
RX   MEDLINE=92306611; PubMed=1819514;
RA   Sikorski R.S., Michaud W.A., Wootton J.C., Boguski M.S., Connelly C.,
RA   Hieter P.A.;
RT   "TPR proteins as essential components of the yeast cell cycle.";
RL   Cold Spring Harb. Symp. Quant. Biol. 56:663-673(1991).
RN   [3]
RP   SUBUNITS.
RX   MEDLINE=95009933; PubMed=7925276;
RA   Lamb J.R., Michaud W.A., Sikorski R.S., Hieter P.A.;
RT   "Cdc16p, Cdc23p and Cdc27p form a complex essential for mitosis.";
RL   EMBO J. 13:4321-4328(1994).
CC   -!- FUNCTION: EXECUTES ESSENTIAL MITOTIC FUNCTIONS NEAR THE
CC       METAPHASE/ANAPHASE TRANSITION.
CC   -!- SUBUNIT: CDC16, CDC23 AND CDC27 FORMS A MACROMOLECULAR COMPLEX.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE CDC27/NUC2 FAMILY.
CC   -!- SIMILARITY: Contains 8 TPR repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X79489; CAA56022.1; ALT_INIT.
DR   EMBL; Z35845; CAA84905.1; -.
DR   PIR; S45825; S45825.
DR   SGD; S0000180; CDC27.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF00515; TPR; 8.
DR   SMART; SM00028; TPR; 5.
KW   Cell division; Cell cycle; Mitosis; Repeat; TPR repeat;
KW   Nuclear protein.
FT   REPEAT      154    187       TPR 1.
FT   REPEAT      472    505       TPR 2.
FT   REPEAT      540    573       TPR 3.
FT   REPEAT      575    607       TPR 4.
FT   REPEAT      608    641       TPR 5.
FT   REPEAT      643    675       TPR 6.
FT   REPEAT      676    709       TPR 7.
FT   REPEAT      711    743       TPR 8.
FT   DOMAIN      358    391       ASN-RICH.
FT   MUTAGEN     613    613       G->D: IN TEMPERATURE SENSITIVE MUTANT.
SQ   SEQUENCE   758 AA;  85436 MW;  8612EA4504327A02 CRC64;

  Query Match            1.3%;  Score 17;  DB 1;  Length 758;
```

```
  Best Local Similarity   100.0%;  Pred. No. 4.8e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        373 NNNNNNNNNNNNNNNNI 389


RESULT 3
PI4K_DICDI
ID   PI4K_DICDI     STANDARD;      PRT;  1093 AA.
AC   P54677;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphatidylinositol 4-kinase (EC 2.7.1.67) (PI4-kinase) (PtdIns-4-
DE   kinase) (PI4K-alpha).
GN   PIKD OR PIK4.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE=96009592; PubMed=7565716;
RA   Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT   "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
RT   discoideum: biological roles of putative mammalian p110 and yeast
RT   Vps34p PI 3-kinase homologs during growth and development.";
RL   Mol. Cell. Biol. 15:5645-5656(1995).
CC   -!- FUNCTION: ACTS ON PHOSPHATIDYLINOSITOL (PI) IN THE FIRST
CC       COMMITTED STEP IN THE PRODUCTION OF THE SECOND MESSENGER
CC       INOSITOL-1,4,5,-TRISPHOSPHATE.
CC   -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol = ADP +
CC       1-phosphatidyl-1D-myo-inositol 4-phosphate.
CC   -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U23479; AAA85725.1; -.
DR   PIR; T18275; T18275.
DR   DictyDb; DD01102; pikD.
DR   InterPro; IPR000403; PI3_PI4_kinase.
DR   Pfam; PF00454; PI3_PI4_kinase; 1.
DR   SMART; SM00146; PI3Kc; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR   PROSITE; PS50290; PI3_4_KINASE_3; 1.
KW   Transferase; Kinase; Multigene family.
FT   DOMAIN      146    156       ASP-RICH.
FT   DOMAIN      184    200       POLY-ASN.
FT   DOMAIN      203    206       POLY-ASN.
FT   DOMAIN      227    233       POLY-ASN.
FT   DOMAIN      277    315       POLY-ASN.
FT   DOMAIN      441    445       POLY-ASP.
FT   DOMAIN      455    466       POLY-THR.
FT   DOMAIN      494    501       POLY-GLY.
FT   DOMAIN      690    696       POLY-THR.
FT   DOMAIN      751    754       POLY-GLN.
FT   DOMAIN      761    772       POLY-THR.
FT   DOMAIN      775    785       POLY-THR.
FT   DOMAIN      833   1093       PI3K/PI4K.
SQ   SEQUENCE   1093 AA;  123017 MW;  8762BC78355AA635 CRC64;

  Query Match              1.3%;  Score 17;  DB 1;  Length 1093;
  Best Local Similarity   100.0%;  Pred. No. 6.8e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        282 NNNNNNNNNNNNNNNNI 298


RESULT 4
PH81_YEAST
ID   PH81_YEAST     STANDARD;      PRT;  1178 AA.
AC   P17442; Q06887;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphate system positive regulatory protein PHO81 (CDK inhibitor
DE   PHO81).
GN   PHO81 OR YGR233C OR G8567.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRF88;
RX   MEDLINE=90245670; PubMed=2186378;
RA   Coche T., Prozzi D., Legrain M., Hilger F., Vandenhaute J.;
RT   "Nucleotide sequence of the PHO81 gene involved in the regulation of
RT   the repressible acid phosphatase gene in Saccharomyces cerevisiae.";
RL   Nucleic Acids Res. 18:2176-2176(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93261428; PubMed=8492812;
RA   Ogawa N., Noguchi K.-I., Yamashita Y., Yasuhara T., Hayashi N.,
RA   Yoshida K., Oshima Y.;
RT   "Promoter analysis of the PHO81 gene encoding a 134 kDa protein
RT   bearing ankyrin repeats in the phosphatase regulon of Saccharomyces
RT   cerevisiae.";
RL   Mol. Gen. Genet. 238:444-454(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=96267763; PubMed=8701610;
RA   van der Aart Q.J.M., Kleine K., Steensma H.Y.;
RT   "Sequence analysis of the 43 kb CRM1-YLM9-PET54-DIE2-SMI1-PHO81-YHB4-
RT   PFK1 region from the right arm of Saccharomyces cerevisiae chromosome
RT   VII.";
RL   Yeast 12:385-390(1996).
RN   [4]
RP   SEQUENCE OF 1-10 FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=93261834; PubMed=8493108;
RA   Creasy C.L., Madden S.L., Bergman L.W.;
RT   "Molecular analysis of the PHO81 gene of Saccharomyces cerevisiae.";
RL   Nucleic Acids Res. 21:1975-1982(1993).
RN   [5]
RP   FUNCTION.
RX   MEDLINE=95025879; PubMed=7939631;
RA   Schneider K.R., Smith R.L., O'Shea E.K.;
RT   "Phosphate-regulated inactivation of the kinase PHO80-PHO85 by the
RT   CDK inhibitor PHO81.";
RL   Science 266:122-126(1994).
CC   -!- FUNCTION: UNDER LOW-PHOSPHATE CONDITION PHO81 INHIBITS THE KINASE
CC       ACTIVITY OF THE PHO80(CYCLIN)-PHO85(KINASE) COMPLEX.
CC   -!- SUBUNIT: ASSOCIATES SPECIFICALLY WITH THE PHO80-PHO85 COMPLEX, AND
CC       MUCH OF THIS INTERACTION IS MEDIATED THROUGH THE PHO80 CYCLIN
CC       SUBUNIT.
CC   -!- SIMILARITY: Contains 6 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   -------------------------------------------------------------------
DR   EMBL; X52482; CAA36726.1; -.
DR   EMBL; D13228; BAA02508.1; -.
DR   EMBL; X87941; CAA61183.1; -.
DR   EMBL; Z73018; CAA97261.1; -.
DR   EMBL; S61041; AAD13922.1; -.
DR   PIR; S57698; S57698.
DR   SGD; S0003465; PHO81.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR004331; SPX.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF03105; SPX; 1.
DR   SMART; SM00248; ANK; 3.
DR   PROSITE; PS50088; ANK_REPEAT; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat; Repeat.
FT   DOMAIN       198    253       ASN/ASP-RICH.
FT   DOMAIN       231    248       POLY-ASN.
FT   REPEAT       423    452       ANK 1.
FT   REPEAT       458    487       ANK 2.
FT   REPEAT       506    535       ANK 3.
FT   REPEAT       556    586       ANK 4.
FT   REPEAT       591    620       ANK 5.
FT   REPEAT       624    653       ANK 6.
FT   CONFLICT     248    248       N -> NN (IN REF. 2).
FT   CONFLICT     728    728       T -> I (IN REF. 2).
FT   CONFLICT     762    762       S -> F (IN REF. 2).
FT   CONFLICT     845    845       D -> H (IN REF. 2).
FT   CONFLICT     873    873       N -> K (IN REF. 2).
FT   CONFLICT     920    920       MISSING (IN REF. 1).
FT   CONFLICT     984    984       A -> V (IN REF. 2).
SQ   SEQUENCE   1178 AA;  134028 MW;  9314EDB94B3F6B7D CRC64;

  Query Match             1.3%;  Score 17;  DB 1;  Length 1178;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-08;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        233 NNNNNNNNNNNNNNNNI 249


RESULT 5
AAC4_DICDI
ID   AAC4_DICDI     STANDARD;      PRT;   317 AA.
AC   P14198;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   AAC-rich mRNA clone PLK330 protein (Fragment).
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90066348; PubMed=2511421;
RA   Shaw D.R., Richter H., Giorda R., Ohmachi T., Ennis H.L.;
RT   "Nucleotide sequences of Dictyostelium discoideum developmentally
RT   regulated cDNAs rich in (AAC) imply proteins that contain clusters of
RT   asparagine, glutamine, or threonine.";
RL   Mol. Gen. Genet. 218:453-459(1989).
CC   -!- DEVELOPMENTAL STAGE: THE CONCENTRATION OF AAC-RICH MRNAS IS LOW
CC       IN DORMANT SPORES AND GROWING CELLS, BUT INCREASES DURING
CC       SPORE-GERMINATION AND MULTICELLULAR DEVELOPMENT.
CC   -!- MISCELLANEOUS: SEVERAL PROTEINS DERIVE FROM AAC-RICH MRNA, WHICH,
CC       DUE TO A FRAMESHIFT ALSO HAVE ACA AND CAA CODONS AND THUS ARE
CC       ASN-, THR- OR GLN-RICH.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; X16523; CAA34530.1; -.
DR   PIR; S05356; S05356.
DR   DictyDb; DD05008; -.
KW   Repeat.
FT   DOMAIN        54     73       ASN-RICH.
FT   DOMAIN       314    317       ASN-RICH.
FT   NON_TER      317    317
SQ   SEQUENCE   317 AA;  35893 MW;  7FB8B5242D8D2CFC CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 2e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db         54 NNNNNNNNNNNNNNNN 69


RESULT 6
MYBH_DICDI
ID   MYBH_DICDI     STANDARD;      PRT;   451 AA.
AC   P34127;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myb-like protein (Fragment).
GN   MYBA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92195700; PubMed=1549373;
RA   Stober-Graesser U., Brydolf B., Bin X., Graesser F., Firtel R.A.,
RA   Lipsick J.S.;
RT   "The Myb DNA-binding domain is highly conserved in Dictyostelium
RT   discoideum.";
RL   Oncogene 7:589-596(1992).
CC   -!- FUNCTION: MAY CONTROL CELLULAR DIFFERENTIATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: Contains 3 Myb-like domains.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-6, MET-110, MET-114
CC       OR MET-123 IS THE INITIATOR.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; Z11534; CAB37862.1; -.
DR   HSSP; P06876; 1MBG.
DR   DictyDb; DD05044; mybA.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   Pfam; PF00249; myb_DNA-binding; 3.
DR   SMART; SM00717; SANT; 3.
DR   PROSITE; PS00037; MYB_1; 3.
DR   PROSITE; PS00334; MYB_2; 3.
DR   PROSITE; PS50090; MYB_3; 3.
KW   Nuclear protein; DNA-binding; Repeat.
FT   DNA_BIND     144    195       MYB 1.
FT   DNA_BIND     196    247       MYB 2.
FT   DNA_BIND     248    298       MYB 3.
FT   DOMAIN        25     61       ASN-RICH.
FT   DOMAIN        89    139       ASN-RICH.
FT   DOMAIN       353    363       PRO-RICH.
FT   DOMAIN        48     61       POLY-ASN.
FT   DOMAIN        62     67       POLY-GLU.
```

```
FT   DOMAIN       68     74       POLY-ASP.
FT   DOMAIN      108    132       POLY-ASN.
FT   DOMAIN      338    349       POLY-THR.
FT   DOMAIN      429    451       POLY-ASN.
FT   NON_TER     451    451
SQ   SEQUENCE   451 AA;  51298 MW;  E199B3B471728F60 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        429 NNNNNNNNNNNNNNNN 444


RESULT 7
CAR3_DICDI
ID   CAR3_DICDI     STANDARD;      PRT;   490 AA.
AC   P35352;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cyclic AMP receptor 3.
GN   CARC OR CAR3.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE=93170666; PubMed=8382181;
RA   Johnson R.L., Saxe C.L. III, Gollop R., Kimmel A.R., Devreotes P.N.;
RT   "Identification and targeted gene disruption of cAR3, a cAMP receptor
RT   subtype expressed during multicellular stages of Dictyostelium
RT   development.";
RL   Genes Dev. 7:273-282(1993).
CC   -!- FUNCTION: RECEPTOR FOR CAMP. COORDINATES THE AGGREGATION
CC       OF INDIVIDUAL CELLS INTO A MULTICELLULAR ORGANISM AND REGULATES
CC       THE EXPRESSION OF A LARGE NUMBER OF DEVELOPMENTALLY REGULATED
CC       GENES. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- DEVELOPMENTAL STAGE: INDUCED AT EARLY AGGREGATION (6 HRS) AND
CC       MAXIMALLY EXPRESSED AT THE MOUND STAGE (9-12HRS), LEVEL OF
CC       EXPRESSION PEAKS AGAIN DURING THE SLUG STAGE (18HRS) AND DECLINES
CC       AT CULMINATION.
CC   -!- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED.
CC   -!- SIMILARITY: BELONGS TO FAMILY 5 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-15 IS THE INITIATOR.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; S55235; AAB25437.1; -.
DR   DictyDb; DD02033; carC.
DR   InterPro; IPR000848; GPCR_cAMP.
DR   InterPro; IPR000832; GPCR_secretin.
DR   PRINTS; PR00247; GPCRCAMP.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation; Multigene family.
FT   DOMAIN        1     23       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     24     43       1 (POTENTIAL).
FT   DOMAIN       44     57       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     58     78       2 (POTENTIAL).
FT   DOMAIN       79     94       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     95    120       3 (POTENTIAL).
FT   DOMAIN      121    131       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    132    150       4 (POTENTIAL).
FT   DOMAIN      151    173       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    174    192       5 (POTENTIAL).
FT   DOMAIN      193    216       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    217    235       6 (POTENTIAL).
FT   DOMAIN      236    246       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    247    271       7 (POTENTIAL).
FT   DOMAIN      272    490       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     204    204       PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT   DOMAIN      303    439       ASN-RICH.
FT   DOMAIN      399    427       POLY-ASN.
SQ   SEQUENCE   490 AA;  56161 MW;  A28BA834D8626153 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 490;
  Best Local Similarity   100.0%;  Pred. No. 3e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        399 NNNNNNNNNNNNNNNN 414


RESULT 8
MOT3_YEAST
ID   MOT3_YEAST     STANDARD;      PRT;   490 AA.
AC   P54785;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein MOT3/HMS1.
GN   MOT3 OR HMS1 OR YMR070W OR YM9916.09.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RA   Madison J., Winston F.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U25279; AAC49982.1; -.
DR   EMBL; Z48952; CAA88795.1; -.
DR   PIR; S52830; S52830.
DR   TRANSFAC; T03448; -.
DR   TRANSFAC; T03500; -.
DR   SGD; S0004674; MOT3.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0003677; F:DNA binding activity; IPI.
DR   GO; GO:0016564; F:transcriptional repressor activity; IDA.
DR   GO; GO:0006350; P:transcription; IGI.
DR   InterPro; IPR007087; Znf_C2H2.
```

```
DR   Pfam; PF00096; zf-C2H2; 2.
DR   SMART; SM00355; ZnF_C2H2; 2.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
KW   Nuclear protein; Zinc-finger; Metal-binding; Repeat.
FT   ZN_FING     346    368       C2H2-TYPE 1.
FT   ZN_FING     374    397       C2H2-TYPE 2.
FT   DOMAIN        8     35       POLY-GLN.
FT   DOMAIN       98    104       POLY-ASN.
FT   DOMAIN      143    157       POLY-ASN.
FT   DOMAIN      173    177       POLY-ALA.
FT   DOMAIN      240    245       POLY-HIS.
FT   DOMAIN      417    420       POLY-SER.
FT   DOMAIN      421    433       POLY-ASN.
FT   DOMAIN      441    450       POLY-ALA.
SQ   SEQUENCE   490 AA;  54382 MW;  4D0DA8DE43F171ED CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 490;
  Best Local Similarity   100.0%;  Pred. No. 3e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        743 NNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        143 NNNNNNNNNNNNNNNI 158


RESULT 9
HSF1_ARATH
ID   HSF1_ARATH     STANDARD;      PRT;   495 AA.
AC   P41151; O23615;
DT   01-FEB-1995 (Rel. 31, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Heat shock factor protein 1 (HSF 1) (Heat shock transcription factor
DE   1) (HSTF 1).
GN   HSF1 OR AT4G17750 OR DL4910C.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95036006; PubMed=7948881;
RA   Huebel A., Schoeffl F.;
RT   "Arabidopsis heat shock factor: isolation and characterization of the
RT   gene and the recombinant protein.";
RL   Plant Mol. Biol. 26:353-362(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=98121113; PubMed=9461215;
RA   Bevan M., Bancroft I., Bent E., Love K., Goodman H.M., Dean C.,
RA   Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA   Ridley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P.,
RA   Wedler H., Wedler E., Wambutt R., Weitzenegger T., Pohl T., Terryn N.,
RA   Gielen J., Villarroel R., De Clercq R., van Montagu M., Lecharny A.,
RA   Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S.,
RA   Kotter P., Entian K.-D., Rieger M., Schaefer M., Funk B.,
RA   Mueller-Auer S., Silvey M., James R., Monfort A., Pons A.,
RA   Puigdomenech P., Douka A., Voukelatou E., Milioni D., Hatzopoulos P.,
RA   Piravandi E., Obermaier B., Hilbert H., Duesterhoeft A., Moores T.,
RA   Jones J.D.G., Eneva T., Palme K., Benes V., Rechmann S., Ansorge W.,
RA   Cooke R., Berger C., Delseny M., Voet M., Volckaert G., Mewes H.-W.,
RA   Klosterman S., Schueller C., Chalwatzis N.;
RT   "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT   Arabidopsis thaliana.";
RL   Nature 391:485-488(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
```

```
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
CC   -!- FUNCTION: DNA-BINDING PROTEIN THAT SPECIFICALLY BINDS HEAT SHOCK
CC       PROMOTER ELEMENTS (HSE) AND ACTIVATES TRANSCRIPTION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: Homotrimer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: EXHIBITS TEMPERATURE-DEPENDENT PHOSPHORYLATION
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X76167; CAA53761.1; -.
DR   EMBL; Z97344; CAB10555.1; -.
DR   EMBL; AL161547; CAB78778.1; -.
DR   PIR; F71447; S52641.
DR   HSSP; P22121; 3HSF.
DR   TRANSFAC; T04394; -.
DR   InterPro; IPR000232; HSF_DNA_bind.
DR   InterPro; IPR002341; HSF_ETS.
DR   Pfam; PF00447; HSF_DNA-bind; 1.
DR   PRINTS; PR00056; HSFDOMAIN.
DR   ProDom; PD001788; HSF_DNA_bind; 1.
```

```
DR   SMART; SM00415; HSF; 1.
DR   PROSITE; PS00434; HSF_DOMAIN; 1.
KW   Transcription regulation; Nuclear protein; DNA-binding; Activator;
KW   Phosphorylation; Heat shock; Multigene family.
FT   DNA_BIND     50    144       BY SIMILARITY.
FT   DOMAIN      150    153       POLY-SER.
FT   DOMAIN      204    207       POLY-GLN.
FT   DOMAIN      454    475       POLY-ASN.
FT   CONFLICT    216    216       L -> M (IN REF. 1).
FT   CONFLICT    285    285       E -> D (IN REF. 1).
FT   CONFLICT    336    336       I -> T (IN REF. 1).
FT   CONFLICT    390    390       F -> Y (IN REF. 1).
FT   CONFLICT    462    473       MISSING (IN REF. 1).
SQ   SEQUENCE   495 AA;  55744 MW;  2F0DE55252B5682A CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 495;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        454 NNNNNNNNNNNNNNNN 469

RESULT 10
RAEP_CANAL
ID   RAEP_CANAL     STANDARD;      PRT;   640 AA.
AC   O93831;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   RAB proteins geranylgeranyltransferase component A (RAB escort
DE   protein) (REP).
GN   MRS6.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO 1060;
RA   Ishii N., Aoki Y., Arisawa M.;
RT   "Molecular cloning of Rab geranylgeranyl transferase escort protein
RT   (REP) homologue from Candida albicans.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS UNPRENYLATED RAB, PRESENTS IT TO THE CATALYTIC
CC       COMPONENT B, AND REMAINS BOUND TO IT AFTER THE GERANYLGERANYL
CC       TRANSFER REACTION. THE COMPONENT A MAY BE REGENERATED BY
CC       TRANSFERRING ITS PRENYLATED RAB TO A PROTEIN ACCEPTOR (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TCD/MRS6 FAMILY OF GDP DISSOCIATION
CC       INHIBITOR.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB021317; BAA36167.1; -.
DR   InterPro; IPR002005; Rab_GDI_REP.
DR   Pfam; PF00996; GDI; 1.
DR   PRINTS; PR00891; RABGDIREP.
KW   GTPase activation.
FT   DOMAIN      419    439       POLY-ASN.
SQ   SEQUENCE   640 AA;  72291 MW;  772BB6C1FF005388 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        419 NNNNNNNNNNNNNNNN 434

RESULT 11
KMHB_DICDI
ID   KMHB_DICDI     STANDARD;      PRT;   732 AA.
AC   P90648;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Myosin heavy chain kinase B (EC 2.7.1.129) (MHCK B).
GN   MHKB OR MHCKB.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE=97277316; PubMed=9115238;
RA   Clancy C.E., Mendoza M.G., Naismith T.V., Kolman M.F., Egelhoff T.T.;
RT   "Identification of a protein kinase from Dictyostelium with homology
RT   to the novel catalytic domain of myosin heavy chain kinase A.";
RL   J. Biol. Chem. 272:11812-11815(1997).
CC   -!- FUNCTION: PHOSPHORYLATES THREONINE IN THE C-TERMINAL TAIL REGION
CC       OF MYOSIN II HEAVY CHAIN. THIS PHOSPHORYLATION IS CRITICAL IN
CC       REGULATING THE ASSEMBLY AND DISASSEMBLY OF MYOSIN II FILAMENT.
CC   -!- CATALYTIC ACTIVITY: ATP + [myosin heavy-chain] = ADP + [myosin
CC       heavy-chain] phosphate.
CC   -!- DOMAIN: CONSISTS OF AN N-TERMINAL DOMAIN, A CENTRAL NONREPETITIVE
CC       CATALYTIC DOMAIN, AND A C-TERMINAL DOMAIN WITH SEVEN WD REPEATS.
CC   -!- SIMILARITY: Contains 7 WD repeats.
CC   -!- SIMILARITY: BELONGS TO THE MHCK / EF-2 PROTEIN KINASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U90946; AAB50136.1; -.
DR   DictyDb; DD01087; mhkB.
DR   InterPro; IPR004166; MHCK_EF2_kinase.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF02816; Alpha_kinase; 1.
DR   Pfam; PF00400; WD40; 7.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   ProDom; PD000018; WD40; 2.
DR   SMART; SM00320; WD40; 7.
DR   PROSITE; PS00678; WD_REPEATS_1; 5.
DR   PROSITE; PS50082; WD_REPEATS_2; 6.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat;
KW   WD repeat.
FT   NP_BIND     298    303       ATP (POTENTIAL).
FT   DOMAIN      355    381       POLY-ASN.
FT   REPEAT      458    486       WD 1.
FT   REPEAT      500    528       WD 2.
FT   REPEAT      540    568       WD 3.
FT   REPEAT      580    608       WD 4.
FT   REPEAT      620    648       WD 5.
FT   REPEAT      660    688       WD 6.
FT   REPEAT      700    730       WD 7.
SQ   SEQUENCE   732 AA;  83167 MW;  A7233C4BD56D4088 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 732;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
```

```
                 ||||||||||||||||
Db         355 NNNNNNNNNNNNNNNN 370

RESULT 12
CIGB_DICDI
ID   CIGB_DICDI     STANDARD;      PRT;   735 AA.
AC   Q94481;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protein cigB (Fragment).
GN   CIGB.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Loomis W.F.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO D.DISCOIDEUM CIGA.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U66528; AAB06791.1; -.
DR   DictyDb; DD01058; cigB.
DR   InterPro; IPR000315; Znf_Bbox.
DR   Pfam; PF00643; zf-B_box; 1.
KW   Repeat.
FT   NON_TER       1      1
FT   DOMAIN      151    155       POLY-ASN.
FT   DOMAIN      251    268       POLY-ASN.
FT   DOMAIN      328    667       TANDEM REPEATS.
SQ   SEQUENCE   735 AA;  83443 MW;  9765013D88DA6C8C CRC64;

  Query Match              1.2%;  Score 16;  DB 1;  Length 735;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         742 NNNNNNNNNNNNNNNN 757
               ||||||||||||||||
Db         253 NNNNNNNNNNNNNNNN 268

RESULT 13
YM38_YEAST
ID   YM38_YEAST     STANDARD;      PRT;   758 AA.
AC   Q03825;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 85.0 kDa protein in HLJ1-SMP2 intergenic region.
GN   YMR164C OR YM8520.13C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
CC   -!- SIMILARITY: Contains 1 LisH domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z49705; CAA89800.1; -.
DR   PIR; S54522; S54522.
DR   SGD; S0004774; MSS11.
DR   GO; GO:0005634; C:nucleus; IC.
DR   GO; GO:0003704; F:specific RNA polymerase II transcription fa. . .; IDA.
DR   GO; GO:0045944; P:positive regulation of transcription from P. . .; IDA.
DR   GO; GO:0007124; P:pseudohyphal growth; IGI.
DR   GO; GO:0005983; P:starch catabolism; IMP.
DR   InterPro; IPR006594; LisH.
DR   SMART; SM00667; LisH; 1.
DR   PROSITE; PS50896; LISH; 1.
KW   Hypothetical protein.
FT   DOMAIN       51     83       LISH.
FT   DOMAIN      290    329       POLY-GLN.
FT   DOMAIN      605    637       POLY-ASN.
FT   DOMAIN      653    656       POLY-SER.
SQ   SEQUENCE   758 AA;  85050 MW;  BA05BFC754D9294B CRC64;

  Query Match              1.2%;  Score 16;  DB 1;  Length 758;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         742 NNNNNNNNNNNNNNNN 757
               ||||||||||||||||
Db         605 NNNNNNNNNNNNNNNN 620

RESULT 14
GPR1_YEAST
ID   GPR1_YEAST     STANDARD;      PRT;   961 AA.
AC   Q12361;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   G protein-coupled receptor GPR1.
GN   GPR1 OR YDL035C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98190081; PubMed=9524122;
RA   Xue Y., Batlle M., Hirsch J.P.;
RT   "GPR1 encodes a putative G protein-coupled receptor that associates
RT   with the Gpa2p Galpha subunit and functions in a Ras-independent
RT   pathway.";
RL   EMBO J. 17:1996-2007(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=97197972; PubMed=9046088;
RA   Saren A.M., Laamanen P., Lejarcegui J.B., Paulin L.;
RT   "The sequence of a 36.7 kb segment on the left arm of chromosome IV
RT   from Saccharomyces cerevisiae reveals 20 non-overlapping open reading
RT   frames (ORFs) including SIT4, FAD1, NAM1, RNA11, SIR2,NAT1,PRP9, ACT2
RT   and MPS1 and 11 new ORFs.";
RL   Yeast 13:65-71(1997).
CC   -!- FUNCTION: Seems to associate with GPA2 and act as G protein-
CC       coupled receptor that senses glucose and controls filamentous
CC       growth. It acts upstream of adenylate cyclase and is required for
```

```
CC       glucose activation of cAMP synthesis in concert with a glucose
CC       phosphorylation-dependent mechanism.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: BELONGS TO FAMILY GPR1/GIT3 OF G-PROTEIN COUPLED
CC       RECEPTORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z74083; CAA98593.1; -.
DR   EMBL; Z71781; CAA96454.1; -.
DR   PIR; S67568; S67568.
DR   SGD; S0002193; GPR1.
DR   GO; GO:0005886; C:plasma membrane; IDA.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; IMP.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IMP.
DR   GO; GO:0007124; P:pseudohyphal growth; IMP.
KW   G-protein coupled receptor; Transmembrane.
FT   TRANSMEM     57     79       POTENTIAL.
FT   TRANSMEM     92    114       POTENTIAL.
FT   TRANSMEM    134    156       POTENTIAL.
FT   TRANSMEM    179    198       POTENTIAL.
FT   TRANSMEM    251    273       POTENTIAL.
FT   TRANSMEM    620    642       POTENTIAL.
FT   DOMAIN      501    547       POLY-ASN.
SQ   SEQUENCE   961 AA;  110708 MW;  9889D857872A4209 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 961;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        515 NNNNNNNNNNNNNNNN 530


RESULT 15
YIN0_YEAST
ID   YIN0_YEAST     STANDARD;      PRT;   964 AA.
AC   P40467;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Putative 108.8 kDa transcriptional regulatory protein in FKH1-STH1
DE   intergenic region.
GN   YIL130W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169870;
RA   Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA   Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA   Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA   Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA   Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL   Nature 387:84-87(1997).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: Contains 1 Zn(2)-Cys(6) fungal-type binuclear cluster
CC       domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z38059; CAA86148.1; -.
DR   PIR; S48404; S48404.
DR   HSSP; P08657; 1CLD.
DR   SGD; S0001392; YIL130W.
DR   InterPro; IPR001138; Fungi_TrN.
DR   Pfam; PF04082; Fungal_trans; 1.
DR   Pfam; PF00172; Zn_clus; 1.
DR   SMART; SM00066; GAL4; 1.
DR   PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
DR   PROSITE; PS50048; ZN2_CY6_FUNGAL_2; 1.
KW   Hypothetical protein; Transcription regulation; DNA-binding;
KW   Nuclear protein; Zinc; Metal-binding.
FT   DNA_BIND     21     47       ZN(2)-CYS(6), FUNGAL-TYPE.
FT   DOMAIN      811    896       ASN-RICH.
SQ   SEQUENCE   964 AA;  108780 MW;  AD5ABE59E4B022CC CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 964;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        853 NNNNNNNNNNNNNNNN 868


RESULT 16
PTP3_DICDI
ID   PTP3_DICDI     STANDARD;      PRT;   989 AA.
AC   P54637;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Protein-tyrosine phosphatase 3 (EC 3.1.3.48) (Protein-tyrosine-
DE   phosphate phosphohydrolase 3).
GN   (PTPC1 OR PTP3) AND (PTPC2 OR PTP3).
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE=96189126; PubMed=8628311;
RA   Gamper M., Howard P.K., Hunter T., Firtel R.A.;
RT   "Multiple roles of the novel protein tyrosine phosphatase PTP3 during
RT   Dictyostelium growth and development.";
RL   Mol. Cell. Biol. 16:2431-2444(1996).
CC   -!- FUNCTION: SEEMS TO DEPHOSPHORYLATE A PROTEIN OF 130 kDa (P130).
CC   -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC       tyrosine + phosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: IN THE ANTERIOR-LIKE AND PRESTALK CELL TYPES.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED AT MODERATE LEVELS DURING GROWTH
CC       AND DEVELOPMENT.
CC   -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC       TYROSINE PHOSPHATASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U38197; AAC47041.1; -.
DR   HSSP; Q06124; 2SHP.
DR   DictyDb; DD01111; ptpC1.
DR   DictyDb; DD0????; ptpC2.
```

```
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_PP.
DR   Pfam; PF00102; Y_phosphatase; 1.
DR   PRINTS; PR00700; PRTYPHPHTASE.
DR   SMART; SM00194; PTPc; 1.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW   Hydrolase.
FT   ACT_SITE     649    649       BY SIMILARITY.
FT   DOMAIN       460    716       PROTEIN-TYROSINE PHOSPHATASE.
FT   DOMAIN        64     71       POLY-ASN.
FT   DOMAIN       109    118       POLY-ASN.
FT   DOMAIN       137    190       POLY-ASN.
FT   DOMAIN       249    257       POLY-SER.
FT   DOMAIN       258    265       POLY-THR.
FT   DOMAIN       286    289       POLY-ASN.
FT   DOMAIN       366    371       POLY-SER.
FT   DOMAIN       787    790       POLY-GLN.
FT   DOMAIN       834    839       POLY-GLN.
FT   DOMAIN       883    892       POLY-GLN.
FT   DOMAIN       906    914       POLY-ASN.
FT   DOMAIN       943    963       POLY-ASN.
SQ   SEQUENCE   989 AA;  109995 MW;  9371105AF80974AF CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 989;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        137 NNNNNNNNNNNNNNNN 152


RESULT 17
SUZ2_DROME
ID   SUZ2_DROME     STANDARD;      PRT;  1365 AA.
AC   P25172;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Suppressor 2 of zeste protein (Protein posterior sex combs).
GN   SU(Z)2.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S;
RX   MEDLINE=91279476; PubMed=2057369;
RA   Brunk B.P., Adler P.N.;
RT   "The sequence of the Drosophila regulatory gene Suppressor two of
RT   zeste.";
RL   Nucleic Acids Res. 19:3149-3149(1991).
CC   -!- FUNCTION: REGULATES EXPRESSION OF THE HOMEOTIC SELECTOR GENES BY
CC       INFLUENCING HIGHER-ORDER CHROMATIN STRUCTURE THROUGH INTERACTION
CC       WITH OTHER PROTEINS.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X56798; CAA40134.1; -.
DR   EMBL; X56799; CAA40135.1; -.
DR   PIR; S14871; S14871.
DR   FlyBase; FBgn0008654; Su(z)2.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT   ZN_FING       35     74       RING-TYPE.
FT   DOMAIN       623    628       POLY-GLN.
FT   DOMAIN      1077   1096       POLY-ASN.
FT   DOMAIN      1241   1251       POLY-SER.
FT   CONFLICT     603    603       MISSING (IN REF. 1; CAA40134).
FT   CONFLICT     785    785       K -> N (IN REF. 1; CAA40134).
FT   CONFLICT     831    831       A -> R (IN REF. 1; CAA40134).
FT   CONFLICT     965    965       MISSING (IN REF. 1; CAA40134).
FT   CONFLICT    1065   1065       D -> E (IN REF. 1; CAA40134).
FT   CONFLICT    1076   1076       MISSING (IN REF. 1; CAA40134).
FT   CONFLICT    1287   1287       A -> P (IN REF. 1; CAA40134).
SQ   SEQUENCE   1365 AA;  146058 MW;  7B4BA0F35B0FA683 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 1365;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db       1077 NNNNNNNNNNNNNNNN 1092


RESULT 18
KYK1_DICDI
ID   KYK1_DICDI     STANDARD;      PRT;  1584 AA.
AC   P18160;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Non-receptor tyrosine kinase spore lysis A (EC 2.7.1.112) (Tyrosine-
DE   protein kinase 1).
GN   PYKA OR SPLA OR DPYK1.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JH10;
RX   MEDLINE=97053827; PubMed=8898241;
RA   Nuckolls G.H., Osherov N., Loomis W.F., Spudich J.A.;
RT   "The Dictyostelium dual-specificity kinase splA is essential for
RT   spore differentiation.";
RL   Development 122:3295-3305(1996).
RN   [2]
RP   SEQUENCE OF 1248-1584 FROM N.A.
RX   MEDLINE=90287147; PubMed=1972546;
RA   Tan J.L., Spudich J.A.;
RT   "Developmentally regulated protein-tyrosine kinase genes in
RT   Dictyostelium discoideum.";
RL   Mol. Cell. Biol. 10:3578-3583(1990).
CC   -!- FUNCTION: ESSENTIAL FOR SPORE DIFFERENTIATION.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT WITH A PEAK
CC       DURING THE MOUND STAGE OF MORPHOGENESIS.
CC   -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
CC   -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; U32174; AAB41125.1; -.
DR   EMBL; M33785; AAA33202.1; -.
DR   PIR; T18276; T18276.
DR   DictyDb; DD03010; pykA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001660; SAM.
DR   InterPro; IPR003877; SPRY_receptor.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00536; SAM; 1.
DR   Pfam; PF00622; SPRY; 3.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00454; SAM; 1.
DR   SMART; SM00449; SPRY; 3.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50105; SAM_DOMAIN; 1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT   DOMAIN      908    972       SAM.
FT   DOMAIN      403    420       POLY-ASN.
FT   DOMAIN      428    435       POLY-THR.
FT   DOMAIN      449    480       POLY-ASN.
FT   DOMAIN      483    491       POLY-ASN.
FT   DOMAIN      494    508       POLY-ASN.
FT   DOMAIN      512    532       POLY-ASN.
FT   DOMAIN      596    600       POLY-ASN.
FT   DOMAIN      808    811       POLY-PHE.
FT   DOMAIN     1026   1029       POLY-SER.
FT   DOMAIN     1195   1210       POLY-ASN.
FT   DOMAIN     1215   1220       POLY-GLN.
FT   DOMAIN     1224   1233       POLY-GLN.
FT   DOMAIN     1266   1274       POLY-PRO.
FT   DOMAIN     1289   1561       PROTEIN KINASE.
FT   NP_BIND    1295   1303       ATP (BY SIMILARITY).
FT   BINDING    1316   1316       ATP (BY SIMILARITY).
FT   ACT_SITE   1417   1417       BY SIMILARITY.
FT   CONFLICT   1248   1248       D -> R (IN REF. 2).
FT   CONFLICT   1435   1435       V -> L (IN REF. 2).
SQ   SEQUENCE   1584 AA;  174304 MW;  5D1589458D8E01E3 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 1584;
  Best Local Similarity   100.0%;  Pred. No. 9e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        449 NNNNNNNNNNNNNNNN 464

RESULT 19
P3K3_DICDI
ID   P3K3_DICDI     STANDARD;      PRT;  1585 AA.
AC   P54675;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphatidylinositol 3-kinase 3 (EC 2.7.1.137) (PI3-kinase)
DE   (PtdIns-3-kinase) (PI3K) (Fragment).
GN   PIKC OR PIK3.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE=96009592; PubMed=7565716;
RA   Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT   "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
RT   discoideum: biological roles of putative mammalian p110 and yeast
RT   Vps34p PI 3-kinase homologs during growth and development.";
RL   Mol. Cell. Biol. 15:5645-5656(1995).
```

```
CC   -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol = ADP +
CC       1-phosphatidyl-1D-myo-inositol 3-phosphate.
CC   -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U23478; AAA85723.1; -.
DR   PIR; T18274; T18274.
DR   DictyDb; DD01101; pikC.
DR   InterPro; IPR000403; PI3_PI4_kinase.
DR   InterPro; IPR002420; PI3K_C2.
DR   InterPro; IPR000341; PI3K_ras_bind.
DR   InterPro; IPR001263; PI3Ka.
DR   Pfam; PF00454; PI3_PI4_kinase; 1.
DR   Pfam; PF00792; PI3K_C2; 1.
DR   Pfam; PF00794; PI3K_rbd; 1.
DR   Pfam; PF00613; PI3Ka; 1.
DR   SMART; SM00142; PI3K_C2; 1.
DR   SMART; SM00144; PI3K_rbd; 1.
DR   SMART; SM00145; PI3Ka; 1.
DR   SMART; SM00146; PI3Kc; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR   PROSITE; PS50290; PI3_4_KINASE_3; 1.
KW   Transferase; Kinase; Multigene family; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       58     84       POLY-ASN.
FT   DOMAIN       98    110       POLY-ASN.
FT   DOMAIN      200    226       POLY-ASN.
FT   DOMAIN      239    254       POLY-ASN.
FT   DOMAIN      345    378       POLY-ASN.
FT   DOMAIN      383    390       POLY-ASN.
FT   DOMAIN      484    488       POLY-SER.
FT   DOMAIN      720    737       POLY-GLN.
FT   DOMAIN     1569   1575       POLY-ASN.
FT   DOMAIN     1221   1484       PI3K/PI4K.
FT   DOMAIN     1510   1539       5 X 5 AA APPROXIMATE REPEATS.
FT   REPEAT     1510   1514       1.
FT   REPEAT     1515   1519       2.
FT   REPEAT     1520   1524       3.
FT   REPEAT     1530   1534       4.
FT   REPEAT     1535   1539       5.
FT   DOMAIN     1547   1560       7 X 2 AA TANDEM REPEATS OF K-E.
SQ   SEQUENCE   1585 AA;  180421 MW;  4689B620D2484961 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 1585;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db         58 NNNNNNNNNNNNNNNN 73

RESULT 20
TAGC_DICDI
ID   TAGC_DICDI     STANDARD;      PRT;  1743 AA.
AC   Q23868;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prestalk-specific protein tagC precursor (EC 3.4.21.-).
GN   TAGC.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=97140317; PubMed=8986798;
RA   Shaulsky G., Escalante R., Loomis W.F.;
RT   "Developmental signal transduction pathways uncovered by genetic
RT   suppressors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).
CC   -!- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGC MAY MEDIATE
CC       INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
CC       S8.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
CC       TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC   -!- SIMILARITY: STRONG, TO TAGB.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U60086; AAB03331.1; -.
DR   PIR; T18279; T18279.
DR   DictyDb; DD02060; tagC.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001140; ABC_TM_transpt.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00664; ABC_membrane; 1.
DR   Pfam; PF00005; ABC_tran; 1.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR   PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; FALSE_NEG.
KW   Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
KW   Signal.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28   1743       PRESTALK-SPECIFIC PROTEIN TAGC.
FT   DOMAIN      316    642       PROTEASE.
FT   DOMAIN     1450   1687       ABC TRANSPORTER.
FT   TRANSMEM    962    982       POTENTIAL.
FT   TRANSMEM   1027   1047       POTENTIAL.
FT   TRANSMEM   1072   1092       POTENTIAL.
FT   TRANSMEM   1157   1177       POTENTIAL.
FT   TRANSMEM   1260   1280       POTENTIAL.
FT   TRANSMEM   1288   1308       POTENTIAL.
FT   ACT_SITE    325    325       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    372    372       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    637    637       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1485   1492       ATP (POTENTIAL).
FT   DOMAIN       42     46       POLY-ASN.
FT   DOMAIN       94    103       POLY-ASN.
FT   DOMAIN      643    646       POLY-ALA.
FT   DOMAIN      733    741       POLY-ASN.
FT   DOMAIN      786    792       POLY-SER.
FT   DOMAIN     1337   1340       POLY-GLU.
FT   DOMAIN     1346   1352       POLY-GLY.
FT   DOMAIN     1353   1357       POLY-ASN.
FT   DOMAIN     1358   1364       POLY-ASP.
FT   DOMAIN     1381   1386       POLY-ASN.
FT   DOMAIN     1707   1729       POLY-ASN.
FT   CARBOHYD    390    390       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    614    614       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    689    689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    735    735       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    741    741       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    776    776       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    832    832       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    887    887       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1251   1251       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1385   1385       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1386   1386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1454   1454       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1704   1704       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1743 AA;  194145 MW;  12DB363E2F729839 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 1743;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db       1714 NNNNNNNNNNNNNNNN 1729


RESULT 21 .
P3K2_DICDI
ID   P3K2_DICDI     STANDARD;      PRT;  1858 AA.
AC   P54674;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Phosphatidylinositol 3-kinase 2 (EC 2.7.1.137) (PI3-kinase)
DE   (PtdIns-3-kinase) (PI3K).
GN   PIKB OR PIK2.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE=96009592; PubMed=7565716;
RA   Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT   "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
RT   discoideum: biological roles of putative mammalian p110 and yeast
RT   Vps34p PI 3-kinase homologs during growth and development.";
RL   Mol. Cell. Biol. 15:5645-5656(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol = ADP +
CC       1-phosphatidyl-1D-myo-inositol 3-phosphate.
CC   -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U23477; AAA85722.1; -.
DR   PIR; T18273; T18273.
DR   DictyDb; DD01100; pikB.
DR   InterPro; IPR000403; PI3_PI4_kinase.
DR   InterPro; IPR002420; PI3K_C2.
DR   InterPro; IPR000341; PI3K_ras_bind.
DR   InterPro; IPR001263; PI3Ka.
DR   Pfam; PF00454; PI3_PI4_kinase; 1.
DR   Pfam; PF00792; PI3K_C2; 1.
DR   Pfam; PF00794; PI3K_rbd; 1.
DR   Pfam; PF00613; PI3Ka; 1.
DR   SMART; SM00142; PI3K_C2; 1.
DR   SMART; SM00144; PI3K_rbd; 1.
DR   SMART; SM00145; PI3Ka; 1.
DR   SMART; SM00146; PI3Kc; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR   PROSITE; PS50290; PI3_4_KINASE_3; 1.
```

```
KW   Transferase; Kinase; Multigene family.
FT   DOMAIN       34     40       POLY-GLY.
FT   DOMAIN      166    172       POLY-SER.
FT   DOMAIN      185    226       POLY-ASN.
FT   DOMAIN      227    235       POLY-THR.
FT   DOMAIN      246    253       POLY-SER.
FT   DOMAIN      259    268       POLY-ASN.
FT   DOMAIN      294    303       POLY-THR.
FT   DOMAIN      361    364       POLY-SER.
FT   DOMAIN      369    384       POLY-SER.
FT   DOMAIN      425    429       POLY-GLN.
FT   DOMAIN      439    444       POLY-THR.
FT   DOMAIN      445    454       POLY-SER.
FT   DOMAIN      562    570       POLY-GLY.
FT   DOMAIN      715    727       POLY-THR.
FT   DOMAIN      982    990       POLY-GLN.
FT   DOMAIN     1015   1049       POLY-ASN.
FT   DOMAIN     1598   1858       PI3K/PI4K.
SQ   SEQUENCE   1858 AA;  203945 MW;  A6C0333D4CDDEA03 CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 1858;
  Best Local Similarity   100.0%;  Pred. No. 1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        185 NNNNNNNNNNNNNNNN 200


RESULT 22
TAGB_DICDI
ID   TAGB_DICDI     STANDARD;      PRT;  1905 AA.
AC   P54683;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Prestalk-specific protein tagB precursor (EC 3.4.21.-).
GN   TAGB.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=95262903; PubMed=7744252;
RA   Shaulsky G., Kuspa A., Loomis W.F.;
RT   "A multidrug resistance transporter/serine protease gene is required
RT   for prestalk specialization in Dictyostelium.";
RL   Genes Dev. 9:1111-1122(1995).
CC   -!- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
CC       INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
CC       S8.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
CC       TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC   -!- SIMILARITY: STRONG, TO TAGC.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U20432; AAA62212.1; -.
DR   PIR; T18267; T18267.
DR   MEROPS; S08.UPW; -.
DR   DictyDb; DD02059; tagB.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001140; ABC_TM_transpt.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00664; ABC_membrane; 1.
DR   Pfam; PF00005; ABC_tran; 1.
DR   Pfam; PF00082; Peptidase_S8; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR   PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
KW   Signal.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32   1905       PRESTALK-SPECIFIC PROTEIN TAGB.
FT   DOMAIN      378    700       PROTEASE.
FT   DOMAIN     1518   1756       ABC TRANSPORTER.
FT   TRANSMEM   1011   1031       POTENTIAL.
FT   TRANSMEM   1076   1096       POTENTIAL.
FT   TRANSMEM   1121   1141       POTENTIAL.
FT   TRANSMEM   1210   1230       POTENTIAL.
FT   TRANSMEM   1309   1329       POTENTIAL.
FT   TRANSMEM   1332   1352       POTENTIAL.
FT   ACT_SITE    387    387       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    432    432       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    695    695       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1553   1560       ATP (POTENTIAL).
FT   DOMAIN       63     67       POLY-GLN.
FT   DOMAIN       95    104       POLY-ASN.
FT   DOMAIN      107    134       POLY-ASN.
FT   DOMAIN      311    321       POLY-SER.
FT   DOMAIN      833    837       POLY-SER.
FT   DOMAIN      838    844       POLY-GLY.
FT   DOMAIN      871    876       POLY-LEU.
FT   DOMAIN     1012   1015       POLY-ILE.
FT   DOMAIN     1386   1389       POLY-GLU.
FT   DOMAIN     1398   1404       POLY-GLY.
FT   DOMAIN     1445   1450       POLY-ASN.
FT   DOMAIN     1765   1779       POLY-ASN.
FT   DOMAIN     1782   1785       POLY-SER.
FT   DOMAIN     1807   1812       POLY-PRO.
FT   DOMAIN     1813   1860       POLY-GLN.
FT   DOMAIN     1872   1878       POLY-PRO.
FT   CARBOHYD    594    594       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    672    672       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    747    747       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    823    823       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1172   1172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1522   1522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1658   1658       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1905 AA;  212518 MW;  B8E223FA8B9AE13C CRC64;

  Query Match             1.2%;  Score 16;  DB 1;  Length 1905;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        107 NNNNNNNNNNNNNNNN 122


RESULT 23
ARP2_PLAFA
ID   ARP2_PLAFA     STANDARD;      PRT;   451 AA.
AC   P13824;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Clustered-asparagine-rich protein (Fragment).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86206015; PubMed=3517875;
RA   Wahlgren M., Aaslund L., Franzen L., Sundvall M., Waahlin B.,
RA   Berzins K., McNicol L.A., Bjoerkman A., Wigzell H., Perlmann P.,
RA   Pettersson U.;
RT   "A Plasmodium falciparum antigen containing clusters of asparagine
RT   residues.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:2677-2681(1986).
CC   -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M13021; AAA29485.1; -.
DR   PIR; A23535; A23535.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS50102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW   RNA-binding; Repeat; Malaria.
FT   NON_TER       1      1
FT   DOMAIN       24    114       RNA-BINDING (RRM) 1.
FT   DOMAIN      350    443       RNA-BINDING (RRM) 2.
FT   DOMAIN      119    128       POLY-ASN.
FT   DOMAIN      175    180       POLY-ASN.
FT   DOMAIN      195    210       POLY-ASN.
FT   DOMAIN      237    245       POLY-ASN.
FT   DOMAIN      264    285       POLY-ASN.
SQ   SEQUENCE   451 AA;  52204 MW;  B686354D85F8C293 CRC64;

  Query Match             1.2%;  Score 15;  DB 1;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNN 756
              |||||||||||||||
Db        271 NNNNNNNNNNNNNNN 285


RESULT 24
YP68_YEAST
ID   YP68_YEAST     STANDARD;      PRT;   612 AA.
AC   Q08925;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical 68.7 kDa protein in PPQ1-MFAL1 intergenic region.
GN   YPL184C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313271; PubMed=9169875;
RA   Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA   Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA   Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA   Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA   Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA   Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA   Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA   Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA   Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA   Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA   Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA   Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA   Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA   Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA   Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA   Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL   Nature 387:103-105(1997).
CC   -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC   -!- SIMILARITY: TO S.POMBE SPCC737.01C AND NRD1.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z73540; CAA97894.1; -.
DR   PIR; S65196; S65196.
DR   SGD; S0006105; YPL184C.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 4.
DR   SMART; SM00360; RRM; 4.
DR   PROSITE; PS50102; RRM; 4.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Hypothetical protein; RNA-binding; Repeat.
FT   DOMAIN      201    274       RNA-BINDING (RRM) 1.
FT   DOMAIN      292    379       RNA-BINDING (RRM) 2.
FT   DOMAIN      431    504       RNA-BINDING (RRM) 3.
FT   DOMAIN      522    602       RNA-BINDING (RRM) 4.
FT   DOMAIN        6     19       POLY-ASN.
FT   DOMAIN       22     28       POLY-ASN.
FT   DOMAIN       45     48       POLY-SER.
FT   DOMAIN      413    419       POLY-ALA.
SQ   SEQUENCE   612 AA;  68663 MW;  A1B2EB381C0BBC8E CRC64;

  Query Match             1.2%;  Score 15;  DB 1;  Length 612;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        744 NNNNNNNNNNNNNNI 758
              |||||||||||||||
Db          6 NNNNNNNNNNNNNNI 20


RESULT 25
CCR4_YEAST
ID   CCR4_YEAST     STANDARD;      PRT;   837 AA.
AC   P31384;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Glucose-repressible alcohol dehydrogenase transcriptional effector
DE   (Carbon catabolite repressor protein 4).
GN   CCR4 OR YAL021C OR FUN27.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93093455; PubMed=1459446;
RA   Malvar T., Biron R.W., Kaback D.B., Denis C.L.;
RT   "The CCR4 protein from Saccharomyces cerevisiae contains a
RT   leucine-rich repeat region which is required for its control of ADH2
RT   gene expression.";
RL   Genetics 132:951-962(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=93209532; PubMed=8458570;
```

```
RA   Ouellette B.F.F., Clark M.W., Keng T., Storms R.K., Zhong W.W.,
RA   Zeng B., Fortin N., Delaney S., Barton A.B., Kaback D.B., Bussey H.;
RT   "Sequencing of chromosome I from Saccharomyces cerevisiae: analysis
RT   of a 32 kb region between the LTE1 and SPO7 genes.";
RL   Genome 36:32-42(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94193531; PubMed=8144453;
RA   Barton A.B., Kaback D.B.;
RT   "Molecular cloning of chromosome I DNA from Saccharomyces cerevisiae:
RT   analysis of the genes in the FUN38-MAK16-SPO7 region.";
RL   J. Bacteriol. 176:1872-1880(1994).
CC   -!- FUNCTION: CONTROL OF ADH2 GENE EXPRESSION. IT IS REQUIRED FOR THE
CC       EXPRESSION OF GENES INVOLVED IN NONFERMENTATIVE GROWTH AND IT
CC       MEDIATES OR IS REQUIRED FOR THE ACTION OF THE SPT6 & SPT10 GENES.
CC   -!- MISCELLANEOUS: THE 169 CARBOXYL-TERMINAL RESIDUES ARE IMPORTANT
CC       FOR CCR4 FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE CCR4/NOCTURIN FAMILY.
CC   -!- SIMILARITY: Contains 5 leucine-rich (LRR) repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; S50459; AAB24455.1; -.
DR   EMBL; L05146; AAC04936.1; -.
DR   PIR; S36713; S36713.
DR   TRANSFAC; T03228; -.
DR   SGD; S0000019; CCR4.
DR   GO; GO:0030015; C:CCR4-NOT core complex; IPI.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0000175; F:3'-5' exoribonuclease activity; IDA.
DR   GO; GO:0000289; P:poly(A) tail shortening; IDA.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; IPI.
DR   InterPro; IPR005135; Exo_endo_phos.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF03372; Exo_endo_phos; 1.
DR   Pfam; PF00560; LRR; 3.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00369; LRR_TYP; 1.
KW   Transcription regulation; Repressor; Leucine-rich repeat; Repeat.
FT   DOMAIN       15     24       POLY-GLN.
FT   DOMAIN       77    103       ASN-RICH.
FT   DOMAIN       89    103       POLY-ASN.
FT   DOMAIN      190    206       GLN-RICH.
FT   REPEAT      356    379       LRR 1.
FT   REPEAT      380    402       LRR 2.
FT   REPEAT      403    418       LRR 3.
FT   REPEAT      419    444       LRR 4.
FT   REPEAT      445    467       LRR 5.
FT   CONFLICT    544    544       L -> I (IN REF. 2).
FT   CONFLICT    803    803       E -> V (IN REF. 2).
SQ   SEQUENCE   837 AA;  94699 MW;  B8DCFF24F5BA35A9 CRC64;

  Query Match             1.2%;  Score 15;  DB 1;  Length 837;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNN 756
              |||||||||||||||
Db         89 NNNNNNNNNNNNNNN 103


RESULT 26
YGB4_YEAST
ID   YGB4_YEAST     STANDARD;      PRT;   888 AA.
AC   P25339;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hypothetical regulatory protein in PDR6-PDR1 intergenic region.
GN   YGL014W OR YGL023.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL125-2B;
RX   MEDLINE=91353086; PubMed=1909073;
RA   Chen W., Balzi E., Capieaux E., Goffeau A.;
RT   "The YGL023 gene encodes a putative regulatory protein.";
RL   Yeast 7:309-312(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL125-2B;
RX   MEDLINE=91353083; PubMed=1882553;
RA   Chen W., Balzi E., Capieaux E., Choder M., Goffeau A.;
RT   "The DNA sequencing of the 17 kb HindIII fragment spanning the LEU1
RT   and ATE1 loci on chromosome VII from Saccharomyces cerevisiae reveals
RT   the PDR6 gene, a new member of the genetic network controlling
RT   pleiotropic drug resistance.";
RL   Yeast 7:287-299(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Hebling U., Hofmann B., Delius H.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Is not essential for haploid growth, but may affect
CC       diploid formation.
CC   -!- SIMILARITY: BELONGS TO THE PUMILIO/MPT5 FAMILY.
CC   -!- SIMILARITY: SOME, TO YEAST YJR091C.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; S57889; AAB19616.1; -.
DR   EMBL; S58126; AAD13898.1; -.
DR   EMBL; Z72536; CAA96714.1; -.
DR   PIR; S64016; S64016.
DR   SGD; S0002982; PUF4.
DR   GO; GO:0007569; P:cell aging; IGI.
DR   GO; GO:0000288; P:mRNA catabolism, deadenylation-dependent; IGI.
DR   GO; GO:0006605; P:protein targeting; IMP.
DR   InterPro; IPR001313; Pumilio/Puf.
DR   Pfam; PF00806; PUF; 8.
DR   SMART; SM00025; Pumilio; 8.
KW   Hypothetical protein; Transcription regulation; Zinc-finger;
KW   Metal-binding; Repeat.
FT   DOMAIN        8     38       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      471    542       ASN-RICH.
FT   DOMAIN      521    535       POLY-ASN (POTENTIAL ACTIVATING DOMAIN).
FT   ZN_FING     724    742       C4-TYPE.
FT   REPEAT      571    583
FT   REPEAT      607    619
FT   REPEAT      643    655
FT   REPEAT      679    691
FT   REPEAT      716    728
FT   REPEAT      752    764
FT   REPEAT      831    843
FT   CONFLICT    595    595       A -> R (IN REF. 1 AND 2).
SQ   SEQUENCE   888 AA;  97797 MW;  659BA1062439F642 CRC64;

  Query Match             1.2%;  Score 15;  DB 1;  Length 888;
  Best Local Similarity   100.0%;  Pred. No. 5e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        742 NNNNNNNNNNNNNNN 756
              |||||||||||||||
Db        521 NNNNNNNNNNNNNNN 535


RESULT 27
FAB1_YEAST
ID   FAB1_YEAST     STANDARD;      PRT;  2278 AA.
AC   P34756;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-phosphatidylinositol-3-phosphate 5-kinase FAB1 (EC 2.7.1.150)
DE   (Phosphatidylinositol-3-phosphate 5-kinase) (Type III PIP kinase).
GN   FAB1 OR YFR019W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95392039; PubMed=7663021;
RA   Yamamoto A., Dewald D.B., Boronenkov I.V., Anderson R.A., Emr S.D.,
RA   Koshland D.;
RT   "Novel PI(4)P 5-kinase homologue, Fab1p, essential for normal vacuole
RT   function and morphology in yeast.";
RL   Mol. Biol. Cell 6:525-539(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=95400292; PubMed=7670463;
RA   Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA   Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA   Yamazaki M., Tashiro H., Eki T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=99035548; PubMed=9811604;
RA   Cooke F.T., Dove S.K., McEwen R.K., Painter G., Holmes A.B.,
RA   Hall M.N., Michell R.H., Parker P.J.;
RT   "The stress-activated phosphatidylinositol 3-phosphate 5-kinase Fab1p
RT   is essential for vacuole function in S. cerevisiae.";
RL   Curr. Biol. 8:1219-1222(1998).
RN   [4]
RP   SIMILARITY TO TCP1/CCT FAMILY.
RX   MEDLINE=95314774; PubMed=7794526;
RA   Waldmann T., Lupas A., Kellermann J., Peters J., Baumeister W.;
RT   "Primary structure of the thermosome from Thermoplasma acidophilum.";
RL   Biol. Chem. Hoppe-Seyler 376:119-126(1995).
CC   -!- FUNCTION: Catalyzes the phosphorylation of phosphatidylinositol-3-
CC       phosphate on the fifth hydroxyl of the myo-inositol ring, to form
CC       phosphatidylinositol-4,5-biphosphate. Required for endocytic-
CC       vacuolar pathway and nuclear migration. The product of the
CC       reaction it catalyzes functions as an important regulator of
CC       vacuole homeostasis perhaps by controlling membrane flux to and/or
CC       from the vacuole.
CC   -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol 4-
CC       monophosphate = ADP + 1-phosphatidyl-1D-myo-inositol 4,5-
CC       bisphosphate.
CC   -!- COFACTOR: Magnesium or Manganese.
CC   -!- SUBCELLULAR LOCATION: VACUOLE-ASSOCIATED (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE PTDINS(4)P-5-KINASE FAMILY.
CC   -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U01017; AAA81360.1; -.
DR   EMBL; D50617; BAA09258.1; -.
DR   PIR; S56274; S56274.
DR   SGD; S0001915; FAB1.
DR   GO; GO:0000285; F:1-phosphatidylinositol-3-phosphate 5-kinase. . .; IDA.
DR   GO; GO:0006644; P:phospholipid metabolism; IGI.
DR   GO; GO:0007033; P:vacuole organization and biogenesis; IMP.
DR   InterPro; IPR002498; PIP5K.
DR   InterPro; IPR000306; Znf_FYVE.
DR   Pfam; PF01363; FYVE; 1.
DR   Pfam; PF01504; PIP5K; 1.
DR   SMART; SM00064; FYVE; 1.
DR   SMART; SM00330; PIPKc; 1.
DR   PROSITE; PS50178; ZF_FYVE; 1.
KW   Transferase; Kinase; Zinc-finger; Zinc.
FT   ZN_FING     240    299       FYVE-TYPE.
FT   DOMAIN      393    397       POLY-PRO.
FT   DOMAIN      571    590       POLY-ASN.
FT   DOMAIN     1808   1811       POLY-THR.
FT   DOMAIN     1891   1897       POLY-GLN.
FT   CONFLICT   2275   2275       R -> W (IN REF. 2).
SQ   SEQUENCE   2278 AA;  257417 MW;  1A0A30E13165DE41 CRC64;

  Query Match             1.2%;  Score 15;  DB 1;  Length 2278;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNN 756
              |||||||||||||||
Db        576 NNNNNNNNNNNNNNN 590


RESULT 28
CTK2_XENLA
ID   CTK2_XENLA     STANDARD;      PRT;   643 AA.
AC   P79955;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Carboxy-terminal kinesin 2 (XCTK2).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=97201465; PubMed=9049251;
RA   Walczak C.E., Verma S., Mitchison T.J.;
RT   "XCTK2: a kinesin-related protein that promotes mitotic spindle
RT   assembly in Xenopus laevis egg extracts.";
RL   J. Cell Biol. 136:859-870(1997).
CC   -!- FUNCTION: PROMOTES MITOTIC SPINDLE ASSEMBLY.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS; A SMALL GLOBULAR
CC       N-TERMINAL, A CENTRAL ALPHA-HELICAL COILED COIL AND A LARGE
CC       GLOBULAR C-TERMINAL WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY
CC       (IT HYDROLYZES ATP AND BINDS MICROTUBULES).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. NCD
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U82809; AAB40402.1; -.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
```

```
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    116       GLOBULAR.
FT   DOMAIN      117    296       COILED COIL (POTENTIAL).
FT   DOMAIN      291    564       KINESIN-MOTOR.
FT   NP_BIND     386    393       ATP (POTENTIAL).
SQ   SEQUENCE   643 AA;  71948 MW;  D3D92D08B88E5057 CRC64;

  Query Match             1.1%;  Score 14;  DB 1;  Length 643;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        383 FAYGQTGSGKTYTM 396


RESULT 29
NCD_DROME
ID   NCD_DROME      STANDARD;      PRT;   700 AA.
AC   P20480; Q9VAG8;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Claret segregational protein.
GN   NCD OR CA(ND) OR CG7831.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R, and Canton-S; TISSUE=Ovary;
RX   MEDLINE=90231469; PubMed=1691829;
RA   Endow S.A., Henikoff S., Soler-Niedziela L.;
RT   "Mediation of meiotic and early mitotic chromosome segregation in
RT   Drosophila by a protein related to kinesin.";
RL   Nature 345:81-83(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 16-700 FROM N.A.
RX   MEDLINE=90275618; PubMed=2140958;
RA   McDonald H.B., Goldstein L.S.B.;
RT   "Identification and characterization of a gene encoding a
RT   kinesin-like protein in Drosophila.";
RL   Cell 61:991-1000(1990).
RN   [5]
RP   MOTOR DIRECTIONALITY.
RX   MEDLINE=91043032; PubMed=2146510;
RA   Walker R.A., Salmon E.D., Endow S.A.;
RT   "The Drosophila claret segregation protein is a minus-end directed
RT   motor molecule.";
RL   Nature 347:780-782(1990).
RN   [6]
RP   CHARACTERIZATION.
RX   MEDLINE=94155838; PubMed=8112290;
RA   Lockhart A., Cross R.A.;
RT   "Origins of reversed directionality in the ncd molecular motor.";
RL   EMBO J. 13:751-757(1994).
RN   [7]
RP   MUTANT ALLELE NCD(D).
RX   MEDLINE=91122049; PubMed=1825056;
RA   Komma D.J., Horne A.S., Endow S.A.;
RT   "Separation of meiotic and mitotic effects of claret
RT   non-disjunctional on chromosome segregation in Drosophila.";
RL   EMBO J. 10:419-424(1991).
RN   [8]
RP   CHARACTERIZATION OF MUTANT ALLELE NCD(D).
RX   MEDLINE=96283629; PubMed=8670831;
RA   Moore J.D., Song H., Endow S.A.;
RT   "A point mutation in the microtubule binding region of the Ncd motor
RT   protein reduces motor velocity.";
RL   EMBO J. 15:3306-3314(1996).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 335-700.
RX   MEDLINE=96195067; PubMed=8606780;
RA   Sablin E.P., Kull F.J., Cooke R., Vale R.D., Fletterick R.J.;
RT   "Crystal structure of the motor domain of the kinesin-related motor
RT   ncd.";
RL   Nature 380:555-559(1996).
CC   -!- FUNCTION: NCD IS REQUIRED FOR NORMAL CHROMOSOMAL SEGREGATION IN
CC       MEIOSIS, IN FEMALES, AND IN EARLY MITOTIC DIVISIONS OF THE EMBRYO.
CC       THE NCD MOTOR ACTIVITY IS DIRECTED TOWARD THE MICROTUBULE'S MINUS
CC       END.
CC   -!- MISCELLANEOUS: NCD(D) IS A MUTANT ALLELE THAT SHOWS ABNORMAL
CC       CHROMOSOMAL SEGREGATION.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. NCD
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X52814; CAA36998.1; -.
DR   EMBL; M33932; AAA28716.1; -.
DR   EMBL; AE003771; AAF56942.1; -.
DR   EMBL; AY058596; AAL13825.1; -.
DR   EMBL; X57475; CAA40713.1; -.
DR   PIR; S09748; S09748.
DR   PDB; 1CZ7; 05-NOV-99.
DR   PDB; 2NCD; 09-JUN-99.
DR   FlyBase; FBgn0002924; ncd.
DR   GO; GO:0007100; P:centrosome separation; IGI.
DR   GO; GO:0000212; P:meiotic spindle assembly; NAS.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW   Meiosis; Mitosis; 3D-structure.
FT   DOMAIN      196    346       COILED COIL (POTENTIAL).
FT   DOMAIN      347    700       KINESIN-MOTOR (BY SIMILARITY).
FT   NP_BIND     434    441       ATP (BY SIMILARITY).
FT   MUTAGEN     556    556       V->F: IN NCD(D); REDUCES MOTOR VELOCITY.
FT   CONFLICT    697    697       S -> N (IN REF. 1).
SQ   SEQUENCE   700 AA;  77473 MW;  ADE043CBCE7FD561 CRC64;

  Query Match             1.1%;  Score 14;  DB 1;  Length 700;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        431 FAYGQTGSGKTYTM 444

RESULT 30
KCM1_XENLA
ID   KCM1_XENLA     STANDARD;      PRT;   730 AA.
AC   Q91636;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Kinesin central motor 1 (XKCM1).
GN   KCM1.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=96140638; PubMed=8548824;
RA   Walczak C.E., Mitchison T.J., Desai A.;
RT   "XKCM1: a Xenopus kinesin-related protein that regulates microtubule
RT   dynamics during mitotic spindle assembly.";
RL   Cell 84:37-47(1996).
RN   [2]
RP   REVISION TO 516.
RC   TISSUE=Ovary;
RA   Walczak C.E.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REGULATES MICROTUBULE DYNAMICS DURING MITOTIC SPINDLE
CC       ASSEMBLY. REQUIRED FOR BOTH ESTABLISHMENT AND MAINTENANCE OF
CC       MITOTIC SPINDLES. MAY ACT DIRECTLY ON MICROTUBULES, CAUSING A
CC       DESTABILIZATION AND EVENTUAL DEPOLYMERIZATION OF THE MICROTUBULE.
CC   -!- SUBUNIT: Homodimer (Potential).
CC   -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
CC       SUBFAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U36485; AAC59743.2; -.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil.
FT   DOMAIN        1    256       GLOBULAR (POTENTIAL).
FT   DOMAIN      257    598       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      599    730       COILED COIL (POTENTIAL).
FT   NP_BIND     352    359       ATP (POTENTIAL).
SQ   SEQUENCE   730 AA;  82585 MW;  25C31C187E491523 CRC64;

  Query Match             1.1%;  Score 14;  DB 1;  Length 730;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 INRSLLALKECIRA 276
              ||||||||||||||
Db        520 INRSLLALKECIRA 533

RESULT 31
ATK2_ARATH
ID   ATK2_ARATH     STANDARD;      PRT;   745 AA.
AC   P46864; Q9T047;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin 2 (Kinesin-like protein B).
GN   ATK2 OR KATB OR AT4G27180 OR T24A18.130.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=94355659; PubMed=8075402;
RA   Mitsui H., Nakatani K., Yamaguchi-Shinozaki K., Shinozaki K.,
RA   Nishikawa K., Takahashi H.;
RT   "Sequencing and characterization of the kinesin-related genes katB
RT   and katC of Arabidopsis thaliana.";
RL   Plant Mol. Biol. 25:865-876(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
```

```
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
CC   -!- FUNCTION: POSSIBLE ROLE IN MITOSIS.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS; A SMALL GLOBULAR
CC       N-TERMINAL, A CENTRAL ALPHA-HELICAL COILED COIL AND A LARGE
CC       GLOBULAR C-TERMINAL WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY
CC       (IT HYDROLYZES ATP AND BINDS MICROTUBULES).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. NCD
CC       SUBFAMILY.
CC   -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC       gene model prediction.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D21137; BAA04673.1; -.
DR   EMBL; AL035680; CAB38848.1; ALT_SEQ.
DR   EMBL; AL161566; CAB79573.1; ALT_SEQ.
DR   PIR; T06048; T06048.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil;
KW   Multigene family.
FT   DOMAIN        1     35       GLOBULAR.
FT   DOMAIN       36    384       COILED COIL.
FT   DOMAIN      385    718       KINESIN-MOTOR (BY SIMILARITY).
FT   NP_BIND     472    479       ATP (POTENTIAL).
SQ   SEQUENCE   745 AA;  84359 MW;  6DB352FDF4FD7BC1 CRC64;
```

```
  Query Match             1.1%;  Score 14;  DB 1;  Length 745;
  Best Local Similarity   100.0%;  Pred. No. 4e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        469 FAYGQTGSGKTYTM 482


RESULT 32
ATK3_ARATH
ID   ATK3_ARATH     STANDARD;      PRT;   754 AA.
AC   P46875; Q9FH38;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin 3 (Kinesin-like protein C).
GN   ATK3 OR KATC OR AT5G54670 OR K5F14.1 OR MRB17.18.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=94355659; PubMed=8075402;
RA   Mitsui H., Nakatani K., Yamaguchi-Shinozaki K., Shinozaki K.,
RA   Nishikawa K., Takahashi H.;
RT   "Sequencing and characterization of the kinesin-related genes katB
RT   and katC of Arabidopsis thaliana.";
RL   Plant Mol. Biol. 25:865-876(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20181125; PubMed=10718197;
RA   Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT   features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:31-63(2000).
CC   -!- FUNCTION: POSSIBLE ROLE IN MITOSIS.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS; A SMALL GLOBULAR
CC       N-TERMINAL, A CENTRAL ALPHA-HELICAL COILED COIL AND A LARGE
CC       GLOBULAR C-TERMINAL WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY
CC       (IT HYDROLYZES ATP AND BINDS MICROTUBULES).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. NCD
CC       SUBFAMILY.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC       gene model prediction.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D21138; BAA04674.1; -.
DR   EMBL; AB022214; BAB09933.1; ALT_SEQ.
DR   PIR; S48020; S48020.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Mitosis;
KW   Multigene family.
FT   DOMAIN        1     45       GLOBULAR.
```

```
FT   DOMAIN       46    393       COILED COIL.
FT   DOMAIN      398    727       KINESIN-MOTOR (BY SIMILARITY).
FT   NP_BIND     481    488       ATP (POTENTIAL).
SQ   SEQUENCE   754 AA;  85030 MW;  76091CD5B5D9C531 CRC64;

  Query Match             1.1%;  Score 14;  DB 1;  Length 754;
  Best Local Similarity   100.0%;  Pred. No. 4e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        478 FAYGQTGSGKTYTM 491


RESULT 33
ATK1_ARATH
ID   ATK1_ARATH     STANDARD;      PRT;   793 AA.
AC   Q07970;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Kinesin 1 (Kinesin-like protein A).
GN   ATK1 OR KATA OR AT4G21270 OR F7J7.210.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=93261419; PubMed=8492804;
RA   Mitsui H., Yamaguchi-Shinozaki K., Shinozaki K., Nishikawa K.,
RA   Takahashi H.;
RT   "Identification of a gene family (kat) encoding kinesin-like proteins
RT   in Arabidopsis thaliana and the characterization of secondary
RT   structure of KatA.";
RL   Mol. Gen. Genet. 238:362-368(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [3]
RP   FUNCTION.
RX   PubMed=12112142;
RA   Marcus A.I., Ambrose J.C., Blickley L., Hancock W.O., Cyr R.J.;
RT   "Arabidopsis thaliana protein, ATK1, is a minus-end directed kinesin
RT   that exhibits non-processive movement.";
RL   Cell Motil. Cytoskeleton 52:144-150(2002).
RN   [4]
RP   FUNCTION.
RX   PubMed=11973272;
RA   Chen C., Marcus A., Li W., Hu Y., Calzada J.-P., Grossniklaus U.,
RA   Cyr R.J., Ma H.;
RT   "The Arabidopsis ATK1 gene is required for spindle morphogenesis in
RT   male meiosis.";
RL   Development 129:2401-2409(2002).
CC   -!- FUNCTION: Kinesin that supports microtubule movement in an ATP-
CC       dependent manner and has a minus-end directed polarity. Plays a
CC       crucial role in spindle morphogenesis in male Arabidopsis meiosis.
CC       In ATK1 mutants, male meiosis is defective, producing an abnormal
CC       number of microspores of variable sizes.
CC   -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS; A SMALL GLOBULAR
CC       N-TERMINAL, A CENTRAL ALPHA-HELICAL COILED COIL AND A LARGE
CC       GLOBULAR C-TERMINAL WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY
CC       (IT HYDROLYZES ATP AND BINDS MICROTUBULES).
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. NCD
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D11371; BAA01972.1; -.
DR   EMBL; AL021960; CAA17546.1; -.
DR   EMBL; AL161554; CAB79127.1; -.
DR   PIR; S34830; S34830.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Meiosis;
KW   Multigene family.
FT   DOMAIN        1     69       GLOBULAR.
FT   DOMAIN       70    428       COILED COIL.
FT   DOMAIN      429    778       KINESIN-MOTOR (BY SIMILARITY).
FT   NP_BIND     516    523       ATP (POTENTIAL).
SQ   SEQUENCE   793 AA;  89047 MW;  85A195206D825C4E CRC64;

  Query Match             1.1%;  Score 14;  DB 1;  Length 793;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          92 FAYGQTGSGKTYTM 105
               ||||||||||||||
Db         513 FAYGQTGSGKTYTM 526


RESULT 34
KLP1_SCHPO
ID   KLP1_SCHPO     STANDARD;      PRT;   832 AA.
AC   Q92376; O42669;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein 1.
GN   PKL1 OR SPAC3A11.14C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=97053988; PubMed=8898367;
RA   Pidoux A.L., Ledizet M., Cande W.Z.;
RT   "Fission yeast pkl1 is a kinesin-related protein involved in mitotic
RT   spindle function.";
RL   Mol. Biol. Cell 7:1639-1655(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: MICROTUBULE-DEPENDENT MOTOR THAT IS INVOLVED IN
CC       MICROTUBULE ORGANIZATION IN THE MITOTIC SPINDLE.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. NCD
CC       SUBFAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U63916; AAB88235.1; -.
DR   EMBL; Z99296; CAB16597.1; -.
DR   EMBL; Z99260; CAB16389.1; -.
DR   PIR; T38749; T38749.
DR   HSSP; P17119; 3KAR.
DR   GeneDB_SPombe; SPAC3A11.14c; -.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil;
KW   Nuclear protein.
FT   DOMAIN      193    266       COILED COIL (POTENTIAL).
FT   DOMAIN      284    305       COILED COIL (POTENTIAL).
FT   DOMAIN      344    481       COILED COIL (POTENTIAL).
FT   DOMAIN      482    832       KINESIN-MOTOR (BY SIMILARITY).
FT   NP_BIND     575    582       ATP (POTENTIAL).
FT   CONFLICT    343    343       E -> D (IN REF. 1).
SQ   SEQUENCE   832 AA;  96336 MW;  870CF74224B9F3E4 CRC64;

  Query Match             1.1%;  Score 14;  DB 1;  Length 832;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          92 FAYGQTGSGKTYTM 105
               ||||||||||||||
Db         572 FAYGQTGSGKTYTM 585


RESULT 35
CYAG_DICDI
ID   CYAG_DICDI     STANDARD;      PRT;   858 AA.
AC   Q03101;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Adenylate cyclase, germination specific (EC 4.6.1.1) (ATP
DE   pyrophosphate-lyase) (Adenylyl cyclase).
GN   ACGA OR ACG.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92233467; PubMed=1348970;
RA   Pitt G.S., Milona N., Borleis J., Lin K.C., Reed R.R.,
RA   Devreotes P.N.;
RT   "Structurally distinct and stage-specific adenylyl cyclase genes play
RT   different roles in Dictyostelium development.";
RL   Cell 69:305-315(1992).
RN   [2]
RP   3D-STRUCTURE MODELING OF 387-543.
RX   MEDLINE=98054247; PubMed=9391039;
RA   Liu Y., Ruoho A.E., Rao V.D., Hurley J.H.;
RT   "Catalytic mechanism of the adenylyl and guanylyl cyclases: modeling
RT   and mutational analysis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:13414-13419(1997).
CC   -!- FUNCTION: ACG HAS A LARGE EXTRACELLULAR DOMAIN WHICH MAY BE
CC       INVOLVED IN THE RECOGNITION OF AN EXTRACELLULAR SIGNAL PRESENT
CC       DURING GERMINATION, LEADING TO ACTIVATION OR INHIBITION OF CAMP
CC       SYNTHESIS BY THE CYTOPLASMIC DOMAIN.
CC   -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC   -!- ENZYME REGULATION: INSENSITIVE TO GUANINE NUCLEOTIDES.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC   -!- DEVELOPMENTAL STAGE: AFTER FRUITING BODIES HAVE BEEN FORMED AND
CC       DURING GERMINATION.
CC   -!- SIMILARITY: Belongs to the adenylyl cyclase class-4/guanylyl
CC       cyclase family.
CC   -!- SIMILARITY: Contains 1 CHASE domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M87278; AAA33164.1; -.
DR   PDB; 1AWM; 28-JAN-98.
DR   DictyDb; DD04007; acgA.
DR   InterPro; IPR006189; CHASE.
DR   InterPro; IPR001054; G_cyclase.
DR   Pfam; PF03924; CHASE; 1.
DR   Pfam; PF00211; guanylate_cyc; 1.
DR   SMART; SM00044; CYCc; 1.
DR   PROSITE; PS50839; CHASE; 1.
DR   PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR   PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW   Lyase; cAMP biosynthesis; Transmembrane; Germination; Signal-anchor;
KW   3D-structure.
FT   DOMAIN        1     18       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     19     41       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       42    858       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN       86    317       CHASE.
FT   DOMAIN      396    526       GUANYLATE CYCLASE.
FT   DOMAIN      738    848       ASN-RICH.
FT   DOMAIN      835    848       POLY-ASN.
FT   STRAND      391    398
FT   HELIX       403    408
FT   HELIX       415    432
FT   TURN        433    434
FT   STRAND      439    443
FT   TURN        444    445
FT   STRAND      446    451
FT   HELIX       462    473
FT   TURN        474    474
FT   TURN        479    482
FT   STRAND      492    492
FT   STRAND      495    500
FT   STRAND      504    504
FT   STRAND      511    511
FT   HELIX       517    527
FT   STRAND      534    534
FT   HELIX       538    542
SQ   SEQUENCE   858 AA;  98413 MW;  B483FB255289E65D CRC64;

  Query Match             1.1%;  Score 14;  DB 1;  Length 858;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNN 755
              ||||||||||||||
Db        835 NNNNNNNNNNNNNN 848

RESULT 36
KINL_LEICH
ID   KINL_LEICH     STANDARD;      PRT;   955 AA.
AC   P46865;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Kinesin-like protein K39 (Fragment).
GN   KIN.
OS   Leishmania chagasi.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=44271;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MHOM/BR/82 / Isolate BA-2;
RX   MEDLINE=93133867; PubMed=8421715;
RA   Burns J.M. Jr., Shreffler W.G., Benson D.R., Ghalib H.W., Badaro R.,
RA   Reed S.G.;
RT   "Molecular characterization of a kinesin-related antigen of
RT   Leishmania chagasi that detects specific antibody in African and
RT   American visceral leishmaniasis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:775-779(1993).
CC   -!- DEVELOPMENTAL STAGE: PREDOMINANT IN AMASTIGOTES.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L07879; AAA29254.1; -.
DR   PIR; A47334; A47334.
DR   HSSP; P17119; 3KAR.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   Motor protein; Microtubules; ATP-binding; Coiled coil; Repeat.
FT   DOMAIN        1    399       KINESIN-MOTOR (BY SIMILARITY).
FT   DOMAIN      426   >955       COILED COIL (POTENTIAL).
FT   NP_BIND     122    129       ATP (POTENTIAL).
FT   DOMAIN      704   >955       7 X 39 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      704    742       1.
FT   REPEAT      743    781       2.
FT   REPEAT      782    820       3.
FT   REPEAT      821    859       4.
FT   REPEAT      860    898       5.
FT   REPEAT      899    937       6.
FT   REPEAT      938   >955       7 (PARTIAL).
FT   NON_TER     955    955
SQ   SEQUENCE   955 AA;  106168 MW;  8CA76815BE84C6E9 CRC64;

  Query Match             1.1%;  Score 14;  DB 1;  Length 955;
  Best Local Similarity   100.0%;  Pred. No. 5e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        119 FAYGQTGSGKTYTM 132

RESULT 37
RI15_YEAST
ID   RI15_YEAST     STANDARD;      PRT;  1770 AA.
AC   P43565;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Serine/threonine-protein kinase RIM15 (EC 2.7.1.-).
GN   RIM15 OR TAK1 OR YFL033C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=95400292; PubMed=7670463;
RA   Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA   Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA   Yamazaki M., Tashiro H., Eki T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
RN   [2]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=97265402; PubMed=9111339;
```

```
RA   Vidan S., Mitchell A.P.;
RT   "Stimulation of yeast meiotic gene expression by the
RT   glucose-repressible protein kinase Rim15p.";
RL   Mol. Cell. Biol. 17:2688-2697(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Reinders A., Buerckert N., Boller T., Wiemken A., de Virgilio C.;
RT   "Saccharomyes cerevisiae cAMP-dependent protein kinase controls entry
RT   into stationary phase through the Rim15p protein kinase.";
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE ACTIVATION OF A MEIOTIC GENES ACTIVATION
CC       PATHWAY.
CC   -!- PTM: AUTOPHOSPHORYLATED.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       STRONG, TO S.POMBE CEK1.
CC   -!- SIMILARITY: Contains 1 response regulatory domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D50617; BAA09206.1; -.
DR   EMBL; U83459; AAB64088.1; -.
DR   EMBL; AJ001030; CAA04486.1; -.
DR   PIR; S56221; S56221.
DR   SGD; S0001861; RIM15.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IDA.
DR   GO; GO:0040020; P:regulation of meiosis; IMP.
DR   GO; GO:0006950; P:response to stress; IMP.
DR   InterPro; IPR000961; Pkinase_C.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001789; Response_reg.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00072; response_reg; 1.
DR   ProDom; PD000001; Prot_kinase; 2.
DR   ProDom; PD000039; Response_reg; 1.
DR   SMART; SM00448; REC; 1.
DR   SMART; SM00133; S_TK_X; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50110; RESPONSE_REGULATORY; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Phosphorylation; Meiosis.
FT   DOMAIN      794   1254       PROTEIN KINASE.
FT   DOMAIN     1636   1750       RESPONSE REGULATORY.
FT   DOMAIN      343    358       POLY-ASN.
FT   DOMAIN      620    624       POLY-SER.
FT   NP_BIND     800    808       ATP (BY SIMILARITY).
FT   BINDING     823    823       ATP (BY SIMILARITY).
FT   ACT_SITE    918    918       BY SIMILARITY.
FT   DOMAIN      975    980       POLY-ASN.
FT   DOMAIN     1213   1218       POLY-GLU.
FT   DOMAIN     1386   1391       POLY-THR.
SQ   SEQUENCE   1770 AA;  196530 MW;  DC1064825000FAFF CRC64;

  Query Match             1.1%;  Score 14;  DB 1;  Length 1770;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNN 755
              ||||||||||||||
Db        345 NNNNNNNNNNNNNN 358


RESULT 38
YM8D_YEAST
ID   YM8D_YEAST     STANDARD;      PRT;   201 AA.
AC   P38429; Q03501;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 23.0 kDa protein in TPS3-IPP2 intergenic region (ORF2).
GN   YMR263W OR YM8156.05.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Manning A.M., Rosenbloom C.L., Beaudet A.L.;
RL   Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M88172; AAA35222.1; -.
DR   EMBL; Z49260; CAA89246.1; -.
DR   PIR; S54475; S54475.
DR   SGD; S0004876; SAP30.
DR   GO; GO:0000118; C:histone deacetylase complex; IPI.
DR   GO; GO:0004407; F:histone deacetylase activity; IPI.
DR   GO; GO:0006347; P:chromatin silencing at HML and HMR (sensu S. . .; IMP.
DR   GO; GO:0000183; P:chromatin silencing at ribosomal DNA (rDNA); IMP.
DR   GO; GO:0006348; P:chromatin silencing at telomere; IMP.
KW   Hypothetical protein.
FT   DOMAIN       26     58       ASN-RICH.
FT   CONFLICT     31     31       C -> S (IN REF. 1).
FT   CONFLICT     48     48       N -> S (IN REF. 1).
SQ   SEQUENCE   201 AA;  23026 MW;  6C0D2734C50DD72E CRC64;

  Query Match             1.0%;  Score 13;  DB 1;  Length 201;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 NNNNNNNNNNNNS 514
              |||||||||||||
Db         39 NNNNNNNNNNNNS 51


RESULT 39
HKLB_LYCES
ID   HKLB_LYCES     STANDARD;      PRT;   426 AA.
AC   O22300;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein knotted-1 like LET12.
GN   LET12.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; lamiids; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. VFNT Cherry;
RX   MEDLINE=98145476; PubMed=9484482;
RA   Janssen B.J., Williams A., Chen J.J., Mathern J., Hake S., Sinha N.;
RT   "Isolation and characterization of two knotted-like homeobox genes
RT   from tomato.";
RL   Plant Mol. Biol. 36:417-425(1998).
CC   -!- FUNCTION: MAY HAVE A ROLE TO PLAY IN FORMATIVE EVENTS IN OVULE AND
CC       EMBRYO MORPHOGENESIS.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN THE MATURE PLANT.
CC   -!- SIMILARITY: BELONGS TO THE TALE/KNOX HOMEOBOX FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF000142; AAC49918.1; -.
DR   PIR; T04318; T04318.
DR   TRANSFAC; T04055; -.
DR   InterPro; IPR005539; ELK.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR005540; KNOX1.
DR   InterPro; IPR005541; KNOX2.
DR   Pfam; PF03789; ELK; 1.
DR   Pfam; PF03790; KNOX1; 1.
DR   Pfam; PF03791; KNOX2; 1.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Homeobox; Nuclear protein.
FT   DOMAIN       15     24       POLY-GLN.
FT   DOMAIN       69     76       POLY-ALA.
FT   DOMAIN      140    152       POLY-ASN.
FT   DOMAIN      283    287       POLY-ASP.
FT   DOMAIN      325    348       ELK DOMAIN.
FT   DNA_BIND    349    411       HOMEOBOX (TALE-TYPE).
SQ   SEQUENCE   426 AA;  47581 MW;  5B52B9E0A34A86BC CRC64;

  Query Match             1.0%;  Score 13;  DB 1;  Length 426;
  Best Local Similarity   100.0%;  Pred. No. 0.00022;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNN 754
              |||||||||||||
Db        140 NNNNNNNNNNNNN 152


RESULT 40
KLP3_SCHPO
ID   KLP3_SCHPO     STANDARD;      PRT;   554 AA.
AC   Q9US60; Q9US61;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinesin-like protein 3 (Kinesin-related protein 1).
GN   KLP3 OR KRP1 OR SPAC1834.07.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RC   STRAIN=972;
RX   MEDLINE=20109189; PubMed=10641037;
RA   Brazer S.-C.W., Williams H.P., Chappell T.G., Cande W.Z.;
RT   "A fission yeast kinesin affects Golgi membrane recycling.";
RL   Yeast 16:149-166(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=22127603; PubMed=12132578;
RA   Jeong J.W., Rhee D.K., Cho S.Y., Hae K.L., Kim D.U., Won M., Kim H.B.;
RT   "Cloning and characterization of the kinesin-related protein, Krp1p,
RT   in Schizosaccharomyces pombe.";
RL   Mol. Cells 13:389-398(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: CYTOPLASMIC MOTOR THAT COULD PLAY A ROLE IN GOLGI
CC       MEMBRANE RECYCLING.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF154055; AAF14525.1; -.
DR   EMBL; AF156966; AAF22609.1; -.
DR   EMBL; AF247188; AAF81205.1; -.
DR   EMBL; AL157734; CAB75775.1; -.
DR   PIR; T50118; T50118.
DR   HSSP; P33176; 1BG2.
DR   GeneDB_SPombe; SPAC1834.07; -.
DR   InterPro; IPR001752; kinesin_motor.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   ATP-binding; Coiled coil; Motor protein.
FT   DOMAIN        1    332       KINESIN-MOTOR (BY
FT                                SIMILARITY).
FT   DOMAIN      446    473       COILED COIL (POTENTIAL).
FT   NP_BIND      84     91       ATP (POTENTIAL).
FT   NP_BIND     233    240       ATP (POTENTIAL).
FT   CONFLICT     57     57       Q -> L (IN REF. 1).
```

```
SQ   SEQUENCE   554 AA;  61938 MW;  B4533BD095C49CBD CRC64;

  Query Match             1.0%;  Score 13;  DB 1;  Length 554;
  Best Local Similarity   100.0%;  Pred. No. 0.00028;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         93 AYGQTGSGKTYTM 105
              |||||||||||||
Db         82 AYGQTGSGKTYTM 94


Search completed: October  2, 2003, 16:32:52
Job time : 51 secs
```

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         October  2, 2003, 16:21:36 ; Search time 94 Seconds
                                             (without alignments)
                                             2174.892 Million cell updates/sec

Title:          US-10-006-780-2
Perfect score:  1288
Sequence:       1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table:  OLIGO
                Gapop 60.0 , Gapext 60.0

Searched:       1107863 seqs, 158726573 residues

Word size :       0

Total number of hits satisfying chosen parameters:      1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :      A_Geneseq_19Jun03:*
                1:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                2:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                3:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                4:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                5:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                6:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                7:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                8:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                9:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                10:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                11:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                12:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                13:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                14:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                15:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                16:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                17:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                18:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                19:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                20:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                21:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                22:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                23:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
                24:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	17	1.3	190	14	AAR41989	Staphylococcus ent
2	17	1.3	190	14	AAR41990	Staphylococcus ent
3	17	1.3	190	14	AAR41991	Staphylococcus ent
4	17	1.3	190	14	AAR41992	Staphylococcus ent
5	17	1.3	190	14	AAR41993	Staphylococcus ent
6	17	1.3	190	14	AAR41994	Staphylococcus ent
7	17	1.3	190	14	AAR41995	Staphylococcus ent
8	17	1.3	190	14	AAR41996	Staphylococcus ent
9	17	1.3	190	14	AAR41997	Staphylococcus ent
10	17	1.3	190	14	AAR41998	Staphylococcus ent
11	17	1.3	190	14	AAR41999	Staphylococcus ent
12	17	1.3	190	14	AAR42000	Staphylococcus ent
13	17	1.3	190	14	AAR42001	Staphylococcus ent
14	17	1.3	190	14	AAR42002	Staphylococcus ent
15	17	1.3	190	14	AAR42003	Staphylococcus ent
16	17	1.3	190	14	AAR42004	Staphylococcus ent
17	17	1.3	190	14	AAR42005	Staphylococcus ent
18	17	1.3	190	14	AAR42006	Staphylococcus ent
19	17	1.3	190	14	AAR42007	Staphylococcus ent
20	17	1.3	190	14	AAR42008	Staphylococcus ent
21	17	1.3	190	14	AAR42009	Staphylococcus ent
22	17	1.3	190	14	AAR42010	Staphylococcus ent
23	17	1.3	190	14	AAR42011	Staphylococcus ent
24	17	1.3	190	14	AAR42012	Staphylococcus ent
25	17	1.3	190	14	AAR42013	Staphylococcus ent
26	16	1.2	141	21	AAG49727	Arabidopsis thalia
27	16	1.2	150	21	AAG49726	Arabidopsis thalia
28	16	1.2	213	23	ABP73877	Candida albicans e
29	16	1.2	256	13	AAR24796	Sequence of fibron
30	16	1.2	286	22	AAY72761	Dictyostelium disc
31	16	1.2	342	17	AAW03625	Human lutropin-cho
32	16	1.2	346	21	AAG16708	Arabidopsis thalia
33	16	1.2	349	21	AAY82392	C. trachomatis MOM
34	16	1.2	357	23	AAO19019	Human kinesin moto
35	16	1.2	357	23	AAU74558	Human kinesin moto
36	16	1.2	376	21	AAG16707	Arabidopsis thalia
37	16	1.2	378	21	AAG16706	Arabidopsis thalia
38	16	1.2	432	21	AAB18414	Amino acid sequenc
39	16	1.2	477	22	AAB30815	Amino acid sequenc
40	16	1.2	563	23	AAO19018	Human kinesin moto
41	16	1.2	563	23	AAU74557	Human kinesin moto
42	16	1.2	590	12	AAR11019	Apo-lipoprotein E
43	16	1.2	686	21	AAB18293	Plasmodium falcipa
44	16	1.2	700	21	AAY58359	Yeast G protein-co
45	16	1.2	710	22	AAB30814	Amino acid sequenc
46	16	1.2	758	22	AAB30812	Amino acid sequenc
47	16	1.2	798	23	ABP73728	Candida albicans e
48	16	1.2	847	21	AAY58360	Yeast G protein-co
49	16	1.2	953	21	AAY58356	Yeast G protein-co
50	16	1.2	953	21	AAY58357	Yeast G protein-co
51	16	1.2	961	21	AAY58355	Yeast G protein-co
52	16	1.2	964	22	AAB30811	Amino acid sequenc
53	16	1.2	964	23	ABP35696	Fungal ZBC protein
54	16	1.2	969	22	ABB63009	Drosophila melanog
55	16	1.2	1075	22	AAB30804	Amino acid sequenc
56	16	1.2	1311	21	AAY71001	Alternative versio
57	16	1.2	1326	22	ABB61156	Drosophila melanog
58	16	1.2	1331	23	AAE29358	Plasmodium falcipa
59	16	1.2	1368	22	ABB60262	Drosophila melanog
60	16	1.2	1817	21	AAB18301	Plasmodium falcipa
61	16	1.2	1844	21	AAB18250	Plasmodium falcipa
62	16	1.2	2150	23	AAO22566	Wooden leg (WOL) g
63	16	1.2	2184	22	AAE00425	P. falciparum telo
64	16	1.2	2380	21	AAB18315	Plasmodium falcipa
65	15	1.2	20	22	AAO05387	Human polypeptide
66	15	1.2	73	22	ABG59908	Human liver peptid
67	15	1.2	181	22	ABB63786	Drosophila melanog
68	15	1.2	423	22	ABG28368	Novel human diagno
69	15	1.2	626	22	ABB59733	Drosophila melanog
70	15	1.2	794	22	AAE00426	P. falciparum telo
71	15	1.2	1316	21	AAB30505	A calcium-dependen
72	15	1.2	1337	21	AAB30504	A calcium-dependen
73	15	1.2	2539	21	AAB18198	Plasmodium falcipa
74	14	1.1	46	22	AAU29815	Novel human secret
75	14	1.1	64	22	ABG03244	Novel human diagno
76	14	1.1	126	23	ABG69155	Human monoclonal a
77	14	1.1	337	23	ABB81634	Human kinesin moto
78	14	1.1	342	23	ABB81633	Human kinesin moto
79	14	1.1	343	21	AAG31719	Arabidopsis thalia
80	14	1.1	348	22	ABU53208	Human cell cycle-a
81	14	1.1	366	22	ABU53125	Intracellular traf
82	14	1.1	373	21	AAG31718	Arabidopsis thalia

83	14	1.1	375	21	AAG31717	Arabidopsis thalia
84	14	1.1	375	21	AAY70952	Arabidopsis SERK i
85	14	1.1	458	23	ABP73738	Candida albicans e
86	14	1.1	693	24	AAE30108	Drosophila melanog
87	14	1.1	700	22	ABB58943	Drosophila melanog
88	14	1.1	730	21	AAY49949	Xenopus laevis kin
89	14	1.1	955	15	AAR57365	K39 polypeptide of
90	14	1.1	955	17	AAW03691	Leishmania chagasi
91	14	1.1	1245	21	AAB18244	Plasmodium falcipa
92	14	1.1	1460	21	AAG40077	Arabidopsis thalia
93	14	1.1	1462	21	AAG40076	Arabidopsis thalia
94	14	1.1	1518	21	AAG40075	Arabidopsis thalia
95	14	1.1	1604	21	AAG31284	Arabidopsis thalia
96	14	1.1	1606	21	AAG31283	Arabidopsis thalia
97	14	1.1	1662	21	AAG31282	Arabidopsis thalia
98	14	1.1	2010	21	AAB18218	Plasmodium falcipa
99	13	1.0	39	12	AAR12902	CSP-2 peptide from
100	13	1.0	48	12	AAR12903	CSP-2 peptide from
101	13	1.0	108	22	AAO00353	Human polypeptide
102	13	1.0	138	21	AAG11796	Arabidopsis thalia
103	13	1.0	147	21	AAG11795	Arabidopsis thalia
104	13	1.0	282	21	AAG43584	Arabidopsis thalia
105	13	1.0	341	23	ABB80079	Human kinesin moto
106	13	1.0	341	24	ABG72398	Human kinesin prot
107	13	1.0	363	21	AAG43583	Arabidopsis thalia
108	13	1.0	367	21	AAG43582	Arabidopsis thalia
109	13	1.0	397	23	ABP73400	Candida albicans e
110	13	1.0	507	22	ABG05000	Novel human diagno
111	13	1.0	679	22	ABB58212	Drosophila melanog
112	13	1.0	790	21	AAG41923	Arabidopsis thalia
113	13	1.0	793	22	ABB71688	Drosophila melanog
114	13	1.0	794	21	AAG41922	Arabidopsis thalia
115	13	1.0	814	21	AAG41921	Arabidopsis thalia
116	13	1.0	941	23	ABP74093	Candida albicans e
117	13	1.0	1003	22	ABB61405	Drosophila melanog
118	13	1.0	1074	22	ABB61124	Drosophila melanog
119	13	1.0	1151	21	AAY83251	F-box protein Grr1
120	13	1.0	1151	23	AAE28254	Yeast Grr1 protein
121	13	1.0	1212	22	ABB71114	Drosophila melanog
122	13	1.0	1279	23	ABG70787	Human kinesin-rela
123	13	1.0	1279	23	ABB80078	Human kinesin moto
124	13	1.0	1279	24	ABG72397	Human partial kine
125	13	1.0	1893	22	ABB59829	Drosophila melanog
126	13	1.0	2573	21	AAB18234	Plasmodium falcipa
127	12	0.9	40	12	AAR12904	CSP-2 peptide from
128	12	0.9	212	23	ABG93169	S. cerevisiae BAX-
129	12	0.9	246	22	AAB47285	rCP41. Cryptospor
130	12	0.9	263	22	ABB70399	Drosophila melanog
131	12	0.9	306	22	ABB60719	Drosophila melanog
132	12	0.9	306	22	ABB60722	Drosophila melanog
133	12	0.9	362	23	AAE14501	Human mitotic cent
134	12	0.9	387	23	ABP73197	Candida albicans e
135	12	0.9	391	23	AAE14500	Human mitotic cent
136	12	0.9	405	23	AAE14504	Human mitotic cent
137	12	0.9	410	21	AAB56496	Human prostate can
138	12	0.9	427	22	ABB58461	Drosophila melanog
139	12	0.9	427	22	ABB65442	Drosophila melanog
140	12	0.9	430	23	AAE14499	Human mitotic cent
141	12	0.9	434	23	AAE14503	Human mitotic cent
142	12	0.9	473	23	AAE14502	Human mitotic cent
143	12	0.9	511	23	ABP73951	Candida albicans e
144	12	0.9	638	22	ABB71320	Drosophila melanog
145	12	0.9	647	11	AAR06001	Insect receptor po
146	12	0.9	675	17	AAW00136	Drosophila FRU 1 p
147	12	0.9	679	20	AAW88456	Human kinesin-rela
148	12	0.9	690	22	AAM39257	Human polypeptide
149	12	0.9	691	21	AAY32253	Drosophila dissati
150	12	0.9	691	22	ABB64167	Drosophila melanog
151	12	0.9	694	22	AAM41043	Human polypeptide
152	12	0.9	694	22	AAM41044	Human polypeptide
153	12	0.9	723	23	AAE14505	Human mitotic cent
154	12	0.9	725	23	AAE14506	Human mitotic cent
155	12	0.9	728	22	AAM39258	Human polypeptide

156	12	0.9	729	22	ABB59500	Drosophila melanog
157	12	0.9	761	22	ABB71561	Drosophila melanog
158	12	0.9	764	21	AAB18282	Plasmodium falcipa
159	12	0.9	803	22	ABB58104	Drosophila melanog
160	12	0.9	842	17	AAW00137	Drosophila FRU 2 p
161	12	0.9	995	22	ABB62510	Drosophila melanog
162	12	0.9	1023	22	AAB11436	D. discoideum ster
163	12	0.9	1341	23	ABP35703	Fungal ZBC protein
164	12	0.9	1493	22	ABB69228	Drosophila melanog
165	12	0.9	1501	23	AAU76762	Plasmodium falcipa
166	12	0.9	1516	21	AAB18195	Plasmodium falcipa
167	12	0.9	1568	24	ABP70152	Amino acid sequenc
168	12	0.9	2061	22	ABB66938	Drosophila melanog
169	12	0.9	3502	22	ABB58382	Drosophila melanog
170	12	0.9	5533	22	ABB65772	Drosophila melanog
171	12	0.9	5560	22	ABB71160	Drosophila melanog
172	11	0.9	121	21	AAG34475	Arabidopsis thalia
173	11	0.9	133	23	ABP01443	Human ORFX protein
174	11	0.9	137	22	AAO05799	Human polypeptide
175	11	0.9	160	22	AAG74164	Human colon cancer
176	11	0.9	205	23	AAE14609	Human microtubule
177	11	0.9	205	24	ABU08213	Human kinesin moto
178	11	0.9	250	23	AAU93167	Arabidopsis transc
179	11	0.9	278	22	ABB69939	Drosophila melanog
180	11	0.9	306	22	ABB70631	Drosophila melanog
181	11	0.9	314	23	AAU82970	C. albicans MTW1 p
182	11	0.9	338	23	ABB07411	Human HsKip3A moto
183	11	0.9	343	23	AAU77185	Human KinI-3 DNA f
184	11	0.9	348	22	ABG27541	Novel human diagno
185	11	0.9	361	23	ABB08099	Human HSET motor p
186	11	0.9	363	23	AAU77186	Human KinI-3 DNA f
187	11	0.9	369	23	ABB08101	Human HSET motor p
188	11	0.9	370	23	ABB08100	Human HSET motor p
189	11	0.9	373	23	AAU77183	Human KinI-3 DNA f
190	11	0.9	375	23	AAU79591	Human HsKif15, kin
191	11	0.9	375	23	AAE14401	Human HsKif15 prot
192	11	0.9	381	23	AAU77184	Human KinI-3 DNA f
193	11	0.9	383	23	ABB07412	Amino acid sequenc
194	11	0.9	383	24	ABG72053	Human HsKip3 fragm
195	11	0.9	409	22	AAU15082	Protein encoded by
196	11	0.9	409	23	ABP73217	Candida albicans e
197	11	0.9	409	23	AAU79592	Human HsKif15, kin
198	11	0.9	409	23	AAE14402	Human HsKif15 prot
199	11	0.9	461	16	AAR75506	Nicotiana alata ar
200	11	0.9	473	23	ABG70992	Human target prote
201	11	0.9	522	23	ABG70991	Human HsKif4 const
202	11	0.9	526	22	ABB68634	Drosophila melanog
203	11	0.9	561	22	ABB67878	Drosophila melanog
204	11	0.9	589	22	ABB58929	Drosophila melanog
205	11	0.9	601	22	ABB61643	Drosophila melanog
206	11	0.9	673	22	AAM38739	Human polypeptide
207	11	0.9	676	23	ABG93214	S. cerevisiae BAX-
208	11	0.9	722	21	AAB18291	Plasmodium falcipa
209	11	0.9	774	22	ABB64148	Drosophila melanog
210	11	0.9	776	22	AAM40525	Human polypeptide
211	11	0.9	782	22	ABB11348	Human kinesin rela
212	11	0.9	864	23	ABB07410	Human kinesin moto
213	11	0.9	864	24	ABG72052	Human kinesin moto
214	11	0.9	1025	23	ABP52133	Plasmodium falcipa
215	11	0.9	1029	22	ABG20649	Novel human diagno
216	11	0.9	1122	21	AAB18295	Plasmodium falcipa
217	11	0.9	1174	22	ABB61704	Drosophila melanog
218	11	0.9	1230	19	AAW48895	Candida albicans C
219	11	0.9	1232	23	ABG70990	Human HsKif4 prote
220	11	0.9	1234	23	ABG70993	Human HsKif4b prot
221	11	0.9	1269	21	AAY77955	A. thaliana enviro
222	11	0.9	1351	21	AAB18290	Plasmodium falcipa
223	11	0.9	1368	23	AAU77182	Human kinesin moto
224	11	0.9	1373	22	ABB71036	Drosophila melanog
225	11	0.9	1388	23	AAU79590	Human kinesin moto
226	11	0.9	1388	23	AAE14400	Human kinesin supe
227	11	0.9	1388	24	ABR48222	Human bladder canc
228	11	0.9	1848	22	ABB68996	Drosophila melanog

229	11	0.9	1945	22	ABB64947	Drosophila melanog
230	11	0.9	1979	21	AAB18171	Plasmodium falcipa
231	11	0.9	1997	21	AAB18287	Plasmodium falcipa
232	11	0.9	2088	22	ABB60656	Drosophila melanog
233	11	0.9	3973	21	AAB18253	Plasmodium falcipa
234	10	0.8	12	24	ABU14073	Transdominant effe
235	10	0.8	13	23	ABP72103	Beta 1,2-N-acetylg
236	10	0.8	16	22	AAU29011	Vector pMAL-p2 mal
237	10	0.8	16	22	AAB97228	Maltose binding pr
238	10	0.8	24	21	AAB10777	B. subtilis lumazi
239	10	0.8	24	22	AAB70814	MBP/BMP fusion con
240	10	0.8	26	22	AAB70815	MBP/BMP fusion con
241	10	0.8	30	22	AAB70816	MBP/BMP fusion con
242	10	0.8	40	18	AAW35993	Peptide SEQ ID NO:
243	10	0.8	40	22	AAU26046	Peptide sequence o
244	10	0.8	62	22	AAO01380	Human polypeptide
245	10	0.8	67	21	AAY69908	Peptide MSI-1922.
246	10	0.8	91	10	AAP90415	Plasmodium falcipa
247	10	0.8	100	20	AAW73960	Human TPC2/MBP fus
248	10	0.8	100	23	AAE15169	Maltose binding pr
249	10	0.8	108	22	AAO02887	Human polypeptide
250	10	0.8	110	22	ABB64832	Drosophila melanog
251	10	0.8	121	22	AAG64025	Human polypeptide
252	10	0.8	147	21	AAG02949	Human secreted pro
253	10	0.8	224	22	ABG10926	Novel human diagno
254	10	0.8	240	23	ABB49146	Listeria monocytog
255	10	0.8	248	22	ABG10925	Novel human diagno
256	10	0.8	265	23	ABP73329	Candida albicans e
257	10	0.8	274	23	ABG70881	Human kinesin HsKi
258	10	0.8	274	23	ABB79895	Human kinesin moto
259	10	0.8	274	23	ABG32861	Human kinesin HsKi
260	10	0.8	281	22	ABG17932	Novel human diagno
261	10	0.8	314	22	ABG14144	Novel human diagno
262	10	0.8	323	22	ABB60025	Drosophila melanog
263	10	0.8	329	23	AAE17787	Human kinesin supe
264	10	0.8	355	22	ABB68677	Drosophila melanog
265	10	0.8	359	23	ABB84482	Human HsKif16b mot
266	10	0.8	359	23	ABB79530	Human kinesin moto
267	10	0.8	359	23	AAE22526	Human HsKif16b mot
268	10	0.8	368	23	ABG93036	S. cerevisiae BAX-
269	10	0.8	378	22	ABB68211	Drosophila melanog
270	10	0.8	412	21	AAY90266	Protein chimera Q.
271	10	0.8	412	24	ABG73864	L. infantum antige
272	10	0.8	437	22	ABB62782	Drosophila melanog
273	10	0.8	437	22	ABB65444	Drosophila melanog
274	10	0.8	456	22	AAB97389	Rice photsensitivi
275	10	0.8	468	22	AAB97390	Rice photsensitivi
276	10	0.8	488	23	ABP73873	Candida albicans e
277	10	0.8	500	19	AAW64057	Human IL-9 recepto
278	10	0.8	501	19	AAW64056	Human IL-9 recepto
279	10	0.8	501	19	AAW64055	Human IL-9 recepto
280	10	0.8	503	21	AAB63190	Human secreted pro
281	10	0.8	504	21	AAB63189	Gene 5 human secre
282	10	0.8	504	22	ABB63967	Drosophila melanog
283	10	0.8	511	23	AAM51137	Maltose binding pr
284	10	0.8	514	21	AAG31268	Arabidopsis thalia
285	10	0.8	515	21	AAG31267	Arabidopsis thalia
286	10	0.8	521	24	AAO19700	IL-13 modulation m
287	10	0.8	522	23	ABB81964	Human IL-9 recepto
288	10	0.8	530	22	ABB61485	Drosophila melanog
289	10	0.8	535	16	AAR78524	Maltose binding pr
290	10	0.8	537	7	AAP60452	Sequence of the As
291	10	0.8	542	23	AAU74356	Human cytoskeleton
292	10	0.8	544	23	ABP74012	Candida albicans e
293	10	0.8	549	22	ABB68830	Drosophila melanog
294	10	0.8	573	21	AAY82154	Pseudomonas putida
295	10	0.8	575	22	ABB63518	Drosophila melanog
296	10	0.8	580	21	AAY82157	Pseudomonas putida
297	10	0.8	580	22	ABB62524	Drosophila melanog
298	10	0.8	596	23	ABG68928	Plasmodium falcipa
299	10	0.8	609	22	ABB61332	Drosophila melanog
300	10	0.8	610	22	ABB71317	Drosophila melanog
301	10	0.8	624	22	ABB70398	Drosophila melanog

302	10	0.8	633	21	AAG31266	Arabidopsis thalia
303	10	0.8	633	22	AAE01905	Arabidopsis thalia
304	10	0.8	644	22	ABB59973	Drosophila melanog
305	10	0.8	645	22	ABB63654	Drosophila melanog
306	10	0.8	659	17	AAW09294	L-proline-4-hydrox
307	10	0.8	676	22	ABB63098	Drosophila melanog
308	10	0.8	679	22	AAB45693	P. falciparum gcpE
309	10	0.8	690	21	AAY96694	MBP-VT-1 subunit A
310	10	0.8	690	23	AAU77830	MBPNVT1-A fusion p
311	10	0.8	694	21	AAY96695	MBP-VT-2 subunit A
312	10	0.8	694	23	AAU77831	MBPNVT2-A fusion p
313	10	0.8	700	23	ABB92360	Herbicidally activ
314	10	0.8	708	17	AAW06411	Maltose binding pr
315	10	0.8	708	21	AAY96690	MBP-VT-1 subunit A
316	10	0.8	708	23	AAU77826	MBPNVT1-A fusion p
317	10	0.8	711	17	AAW06412	Maltose binding pr
318	10	0.8	711	21	AAY96691	MBP-VT-2 subunit A
319	10	0.8	711	23	AAU77827	MBPNVT2-A fusion p
320	10	0.8	718	22	ABB66061	Drosophila melanog
321	10	0.8	720	22	AAE01346	Arabidopsis thalia
322	10	0.8	732	22	ABB65450	Drosophila melanog
323	10	0.8	740	23	ABG71293	Human oestrogen re
324	10	0.8	740	23	AAM51591	Fused ER protein.
325	10	0.8	750	22	AAB30813	Amino acid sequenc
326	10	0.8	752	21	AAW90883	P. falciparum gcpE
327	10	0.8	757	22	AAU19569	Human diagnostic a
328	10	0.8	757	23	ABP51294	Human MDDT SEQ ID
329	10	0.8	762	23	ABG60124	Human DITHP polype
330	10	0.8	763	21	AAY82155	Pseudomonas putida
331	10	0.8	784	20	AAY06618	Thermomyces lanugi
332	10	0.8	801	22	ABB62287	Drosophila melanog
333	10	0.8	809	23	AAM48930	Maltose binding pr
334	10	0.8	839	23	ABG71292	Human androgen rec
335	10	0.8	839	23	AAG68238	Fused androgen rec
336	10	0.8	931	22	ABB61093	Drosophila melanog
337	10	0.8	953	21	AAY82156	Pseudomonas putida
338	10	0.8	988	22	ABB60007	Drosophila melanog
339	10	0.8	1029	23	AAE17786	Human kinesin supe
340	10	0.8	1038	23	AAM48337	Human kinesin supe
341	10	0.8	1059	22	ABB58331	Drosophila melanog
342	10	0.8	1068	23	ABP35714	Fungal ZBC protein
343	10	0.8	1084	19	AAW59033	B. pallidus DNA po
344	10	0.8	1121	22	ABB72021	Drosophila melanog
345	10	0.8	1172	23	ABG93308	C. albicans BAX-as
346	10	0.8	1198	22	ABB60371	Drosophila melanog
347	10	0.8	1276	19	AAW59034	B. pallidus DNA po
348	10	0.8	1308	21	AAB18167	Plasmodium falcipa
349	10	0.8	1362	23	AAU74840	Human HsKif13a pro
350	10	0.8	1375	23	ABB84481	Human HsKif16b pro
351	10	0.8	1375	23	ABB79531	Human kinesin moto
352	10	0.8	1375	23	AAE22525	Human HsKif16b pro
353	10	0.8	1392	24	AAE32129	Human cytoskeleton
354	10	0.8	1419	23	ABP52134	Plasmodium falcipa
355	10	0.8	1438	21	AAB18270	Plasmodium falcipa
356	10	0.8	1598	23	ABP73929	Candida albicans e
357	10	0.8	1805	23	ABP68930	Human polypeptide
358	10	0.8	1817	21	AAB18255	Plasmodium falcipa
359	10	0.8	1838	22	ABB64225	Drosophila melanog
360	10	0.8	1921	22	ABB62962	Drosophila melanog
361	10	0.8	1926	22	AAG84915	Shrimp white spot
362	10	0.8	1991	22	ABB60651	Drosophila melanog
363	10	0.8	2112	22	ABB60403	Drosophila melanog
364	10	0.8	2183	22	ABG29838	Novel human diagno
365	10	0.8	2435	22	ABB60448	Drosophila melanog
366	10	0.8	2532	22	ABB63465	Drosophila melanog
367	10	0.8	2532	22	ABB67166	Drosophila melanog
368	10	0.8	3614	22	ABB62664	Drosophila melanog
369	10	0.8	4134	20	AAY31946	Plasmodium falcipa
370	10	0.8	4904	22	ABB62249	Drosophila melanog
371	9	0.7	34	23	ABG40545	Human peptide enco
372	9	0.7	39	18	AAW33332	Sequence encoded b
373	9	0.7	52	23	ABP31745	Human structural p
374	9	0.7	57	22	ABG52907	Human liver peptid

375	9	0.7	57	22	ABB38053	Peptide #5559 enco
376	9	0.7	57	22	AAM58688	Human brain expres
377	9	0.7	57	22	AAM71195	Human bone marrow
378	9	0.7	57	22	AAM31476	Peptide #5513 enco
379	9	0.7	57	23	ABG40992	Human peptide enco
380	9	0.7	72	22	ABB69713	Drosophila melanog
381	9	0.7	139	22	ABB69010	Drosophila melanog
382	9	0.7	140	21	AAG09377	Arabidopsis thalia
383	9	0.7	145	22	ABB71063	Drosophila melanog
384	9	0.7	159	21	AAB43261	Human ORFX ORF3025
385	9	0.7	167	21	AAG31470	Arabidopsis thalia
386	9	0.7	176	24	AAE30194	Arabidopsis sp. bp
387	9	0.7	177	20	AAY59708	Secreted protein 7
388	9	0.7	180	24	AAE30195	Arabidopsis sp. bp
389	9	0.7	187	21	AAG45405	Arabidopsis thalia
390	9	0.7	208	22	ABG15503	Novel human diagno
391	9	0.7	213	22	AAU19360	Human G protein-co
392	9	0.7	248	21	AAG60582	Arabidopsis thalia
393	9	0.7	257	21	AAB18340	Plasmodium falcipa
394	9	0.7	272	21	AAY97188	Modified F64L GFP
395	9	0.7	280	22	ABB65783	Drosophila melanog
396	9	0.7	284	22	ABB69768	Drosophila melanog
397	9	0.7	285	22	ABB65204	Drosophila melanog
398	9	0.7	300	21	AAG31469	Arabidopsis thalia
399	9	0.7	313	22	ABB62420	Drosophila melanog
400	9	0.7	314	22	ABB67734	Drosophila melanog
401	9	0.7	320	21	AAG45404	Arabidopsis thalia
402	9	0.7	323	21	AAG45403	Arabidopsis thalia
403	9	0.7	324	24	ABJ19039	Pathogen specific
404	9	0.7	330	23	ABP39157	Staphylococcus epi
405	9	0.7	354	23	ABP73242	Candida albicans e
406	9	0.7	357	23	AAU76958	Novel human kinesi
407	9	0.7	372	23	ABB93414	Herbicidally activ
408	9	0.7	378	22	AAB60643	Arabidopsis thalia
409	9	0.7	398	21	AAG29336	Arabidopsis thalia
410	9	0.7	410	23	ABB55031	Lactococcus lactis
411	9	0.7	412	21	AAY97189	GFP F64L/KGF fusio
412	9	0.7	437	22	AAG64524	P. hybrida pollen-
413	9	0.7	437	22	AAG62618	Petunia zinc finge
414	9	0.7	454	22	AAB30816	Amino acid sequenc
415	9	0.7	454	23	ABP73979	Candida albicans e
416	9	0.7	470	22	ABB59219	Drosophila melanog
417	9	0.7	498	23	ABP73393	Candida albicans e
418	9	0.7	502	22	ABB65558	Drosophila melanog
419	9	0.7	508	21	AAB18187	Plasmodium falcipa
420	9	0.7	526	22	ABB67142	Drosophila melanog
421	9	0.7	548	22	ABB59009	Drosophila melanog
422	9	0.7	567	21	AAB18197	Plasmodium falcipa
423	9	0.7	568	22	ABB58131	Drosophila melanog
424	9	0.7	578	22	ABB59076	Drosophila melanog
425	9	0.7	604	23	ABB05593	Mitotic centromere
426	9	0.7	613	22	ABB65716	Drosophila melanog
427	9	0.7	621	22	ABB58165	Drosophila melanog
428	9	0.7	629	22	AAG70887	C albicans apoptos
429	9	0.7	645	22	ABB61545	Drosophila melanog
430	9	0.7	648	22	ABB63109	Drosophila melanog
431	9	0.7	660	22	ABB68167	Drosophila melanog
432	9	0.7	670	23	ABP73731	Candida albicans e
433	9	0.7	675	24	ABP78737	N. gonorrhoeae ami
434	9	0.7	691	22	ABB57876	Drosophila melanog
435	9	0.7	719	22	AAG70700	S cerevisiae apopt
436	9	0.7	724	22	ABB65778	Drosophila melanog
437	9	0.7	726	21	AAG31118	Arabidopsis thalia
438	9	0.7	729	22	AAG70847	C albicans apoptos
439	9	0.7	740	23	ABP73836	Candida albicans e
440	9	0.7	764	23	ABP73760	Candida albicans e
441	9	0.7	769	23	ABP73280	Candida albicans e
442	9	0.7	807	21	AAB18311	Plasmodium falcipa
443	9	0.7	816	24	AAE30116	Fruit fly fushi ta
444	9	0.7	829	21	AAG31117	Arabidopsis thalia
445	9	0.7	834	21	AAG31116	Arabidopsis thalia
446	9	0.7	844	22	ABB66588	Drosophila melanog
447	9	0.7	856	22	ABB60983	Drosophila melanog
448	9	0.7	856	22	ABB68808	Drosophila melanog
449	9	0.7	858	22	ABB60597	Drosophila melanog
450	9	0.7	860	19	AAW63700	Receptor type tyro
451	9	0.7	885	22	ABB60372	Drosophila melanog
452	9	0.7	887	22	ABB71107	Drosophila melanog
453	9	0.7	898	22	ABU53123	Intracellular traf
454	9	0.7	898	23	AAU76957	Novel human kinesi
455	9	0.7	905	23	AAU76967	Novel human kinesi
456	9	0.7	943	22	ABB62085	Drosophila melanog
457	9	0.7	950	22	ABB71271	Drosophila melanog
458	9	0.7	960	22	ABB58434	Drosophila melanog
459	9	0.7	1003	18	AAW19917	Drosophila virilis
460	9	0.7	1018	22	ABB63718	Drosophila melanog
461	9	0.7	1036	22	AAG85011	Shrimp white spot
462	9	0.7	1094	23	ABP35676	Fungal ZBC protein
463	9	0.7	1140	22	ABB71145	Drosophila melanog
464	9	0.7	1143	23	ABP73952	Candida albicans e
465	9	0.7	1165	23	AAE24372	Fruit fly E93 prog
466	9	0.7	1186	22	ABB63516	Drosophila melanog
467	9	0.7	1193	21	AAB18306	Plasmodium falcipa
468	9	0.7	1195	22	ABB63694	Drosophila melanog
469	9	0.7	1207	22	ABB62248	Drosophila melanog
470	9	0.7	1221	23	AAE24373	Fruit fly E93 prog
471	9	0.7	1245	23	ABP73820	Candida albicans e
472	9	0.7	1418	23	ABG93257	C. albicans BAX-as
473	9	0.7	1419	22	ABB65521	Drosophila melanog
474	9	0.7	1436	21	AAB18199	Plasmodium falcipa
475	9	0.7	1445	23	ABP35623	Fungal ZBC protein
476	9	0.7	1460	22	ABB71977	Drosophila melanog
477	9	0.7	1476	22	ABB58706	Drosophila melanog
478	9	0.7	1518	22	ABB64829	Drosophila melanog
479	9	0.7	1537	22	ABB58058	Drosophila melanog
480	9	0.7	1557	22	ABB62485	Drosophila melanog
481	9	0.7	1557	23	ABG70015	Larval viability a
482	9	0.7	1714	21	AAB18275	Plasmodium falcipa
483	9	0.7	1905	22	ABB64925	Drosophila melanog
484	9	0.7	2037	22	ABB69099	Drosophila melanog
485	9	0.7	2206	21	AAB18254	Plasmodium falcipa
486	9	0.7	2485	21	AAB18172	Plasmodium falcipa
487	9	0.7	2938	15	AAR59923	GAP protein Iral.
488	9	0.7	4498	22	ABB58595	Drosophila melanog
489	8	0.6	10	22	AAG83925	Arabidopsis thalia
490	8	0.6	27	17	AAR94328	Anti-thrombogenic
491	8	0.6	39	21	AAG58570	Arabidopsis thalia
492	8	0.6	39	22	ABG58895	Human liver peptid
493	8	0.6	39	22	ABB43516	Peptide #11022 enc
494	8	0.6	39	22	ABB26476	Protein #8475 enco
495	8	0.6	39	22	AAM64453	Human brain expres
496	8	0.6	39	22	AAM77259	Human bone marrow
497	8	0.6	39	22	AAM21195	Peptide #7629 enco
498	8	0.6	39	22	AAM37410	Peptide #11447 enc
499	8	0.6	39	23	ABG46275	Human peptide enco
500	8	0.6	41	22	ABG55641	Human liver peptid
501	8	0.6	41	22	ABB40366	Peptide #7872 enco
502	8	0.6	41	22	AAM61175	Human brain expres
503	8	0.6	41	22	AAM73890	Human bone marrow
504	8	0.6	41	22	AAM34072	Peptide #8109 enco
505	8	0.6	41	23	ABG43778	Human peptide enco
506	8	0.6	42	22	ABB68447	Drosophila melanog
507	8	0.6	52	22	ABG48060	Human liver peptid
508	8	0.6	52	22	ABB28036	Human peptide #687
509	8	0.6	52	22	ABB33210	Peptide #716 encod
510	8	0.6	52	22	ABB18675	Protein #674 encod
511	8	0.6	52	22	AAM54003	Human brain expres
512	8	0.6	52	22	AAM66393	Human bone marrow
513	8	0.6	52	22	AAM14263	Peptide #697 encod
514	8	0.6	52	22	AAM26673	Peptide #710 encod
515	8	0.6	52	22	AAM01993	Peptide #675 encod
516	8	0.6	52	23	ABG36042	Human peptide enco
517	8	0.6	63	21	AAG54171	Arabidopsis thalia
518	8	0.6	71	22	ABB70696	Drosophila melanog
519	8	0.6	75	22	ABG55613	Human liver peptid
520	8	0.6	75	22	AAM61148	Human brain expres

521	8	0.6	75	22	AAM73862	Human bone marrow
522	8	0.6	75	23	ABG43750	Human peptide enco
523	8	0.6	76	19	AAW72051	HSV-2 strain SB5 C
524	8	0.6	82	21	AAB11545	SEN virus protein
525	8	0.6	89	23	ABP31402	Human ORF375 prote
526	8	0.6	90	18	AAW38201	Rice SCARECROW SRP
527	8	0.6	90	21	AAB28581	Rice SCLo4. Oryza
528	8	0.6	95	22	ABG26992	Novel human diagno
529	8	0.6	102	22	ABG26896	Novel human diagno
530	8	0.6	103	22	ABG28278	Novel human diagno
531	8	0.6	107	22	AAO07842	Human polypeptide
532	8	0.6	114	22	ABB11091	Human nuclear pore
533	8	0.6	117	22	ABB67854	Drosophila melanog
534	8	0.6	117	22	ABG27133	Novel human diagno
535	8	0.6	126	22	ABG03470	Novel human diagno
536	8	0.6	134	21	AAG29946	Arabidopsis thalia
537	8	0.6	134	22	AAO01002	Human polypeptide
538	8	0.6	134	22	AAO12584	Human polypeptide
539	8	0.6	134	24	ABP81236	Arabidopsis thalia
540	8	0.6	137	23	ABJ11269	Yeast selected int
541	8	0.6	137	23	ABG77367	Selected Interacti
542	8	0.6	141	21	AAG41855	Arabidopsis thalia
543	8	0.6	143	22	ABB71001	Drosophila melanog
544	8	0.6	145	21	AAG47273	Arabidopsis thalia
545	8	0.6	146	21	AAY96215	Argemone mexicana
546	8	0.6	146	24	ABR40724	Argemone mexicana
547	8	0.6	147	21	AAG47272	Arabidopsis thalia
548	8	0.6	147	23	ABB83125	Human glutary-CoA
549	8	0.6	151	21	AAG29944	Arabidopsis thalia
550	8	0.6	154	22	ABB70370	Drosophila melanog
551	8	0.6	157	21	AAG23969	Arabidopsis thalia
552	8	0.6	164	19	AAW80406	A secreted protein
553	8	0.6	164	23	ABP61824	Human polypeptide
554	8	0.6	167	22	ABB62385	Drosophila melanog
555	8	0.6	173	21	AAG08442	Arabidopsis thalia
556	8	0.6	173	21	AAG45492	Arabidopsis thalia
557	8	0.6	184	21	AAG51370	Arabidopsis thalia
558	8	0.6	185	23	ABJ11066	Yeast selected int
559	8	0.6	185	23	ABG77195	Selected Interacti
560	8	0.6	186	22	ABB69467	Drosophila melanog
561	8	0.6	193	24	ABP71417	Human UBE2 related
562	8	0.6	195	19	AAW59053	His rFGF10 protein
563	8	0.6	215	18	AAW24049	Rat fibroblast gro
564	8	0.6	216	23	ABJ11078	Yeast selected int
565	8	0.6	216	23	ABG77207	Selected Interacti
566	8	0.6	221	22	ABB64291	Drosophila melanog
567	8	0.6	221	23	ABB06719	Human oestrogen re
568	8	0.6	222	23	ABP73512	Candida albicans e
569	8	0.6	227	21	AAG51369	Arabidopsis thalia
570	8	0.6	228	21	AAG15809	Arabidopsis thalia
571	8	0.6	228	22	ABB62497	Drosophila melanog
572	8	0.6	229	21	AAB43838	Human cancer assoc
573	8	0.6	229	21	AAG17420	Arabidopsis thalia
574	8	0.6	232	21	AAG57272	Arabidopsis thalia
575	8	0.6	234	21	AAG51368	Arabidopsis thalia
576	8	0.6	234	23	ABJ11125	Yeast selected int
577	8	0.6	234	23	ABG77254	Selected Interacti
578	8	0.6	235	21	AAG15808	Arabidopsis thalia
579	8	0.6	239	21	AAG57271	Arabidopsis thalia
580	8	0.6	240	22	ABB70127	Drosophila melanog
581	8	0.6	242	22	ABB69486	Drosophila melanog
582	8	0.6	249	22	ABB60887	Drosophila melanog
583	8	0.6	250	23	AAU93090	Arabidopsis transc
584	8	0.6	250	24	ABR40794	Arabidopsis thalia
585	8	0.6	251	23	ABP73583	Candida albicans e
586	8	0.6	255	21	AAG23968	Arabidopsis thalia
587	8	0.6	255	21	AAG43309	Arabidopsis thalia
588	8	0.6	256	21	AAB43638	Human cancer assoc
589	8	0.6	258	21	AAG15807	Arabidopsis thalia
590	8	0.6	259	21	AAG23967	Arabidopsis thalia
591	8	0.6	259	21	AAG43308	Arabidopsis thalia
592	8	0.6	260	21	AAG17419	Arabidopsis thalia
593	8	0.6	265	22	ABG23353	Novel human diagno
594	8	0.6	265	23	AAU74518	Tomato JOINTLESS p
595	8	0.6	267	21	AAG57270	Arabidopsis thalia
596	8	0.6	272	16	AAR74043	Rice carbonic-anhy
597	8	0.6	275	21	AAG09455	Arabidopsis thalia
598	8	0.6	278	21	AAG18317	Arabidopsis thalia
599	8	0.6	278	21	AAG43007	Arabidopsis thalia
600	8	0.6	280	21	AAG24816	Arabidopsis thalia
601	8	0.6	282	21	AAG18316	Arabidopsis thalia
602	8	0.6	282	21	AAG43006	Arabidopsis thalia
603	8	0.6	282	21	AAG46430	Arabidopsis thalia
604	8	0.6	282	22	AAB48841	Petunia hybrida ZP
605	8	0.6	284	22	ABB59014	Drosophila melanog
606	8	0.6	285	23	AAO15048	Rat 187-2 protein
607	8	0.6	285	23	AAO15049	Rat 187-2 protein
608	8	0.6	285	23	AAO15051	Rat 187-2 protein
609	8	0.6	288	21	AAG24815	Arabidopsis thalia
610	8	0.6	288	21	AAG43307	Arabidopsis thalia
611	8	0.6	290	21	AAG46429	Arabidopsis thalia
612	8	0.6	294	16	AAR75512	P. communis (pear)
613	8	0.6	295	24	ABB82605	Potato plastidial
614	8	0.6	299	22	AAM50137	Human kinesin moto
615	8	0.6	299	23	AAO21658	Protein fragment o
616	8	0.6	300	21	AAB05954	Protein deduced fr
617	8	0.6	301	23	ABP38070	Staphylococcus epi
618	8	0.6	302	23	ABJ10412	Myb-related transc
619	8	0.6	303	21	AAG20322	Arabidopsis thalia
620	8	0.6	304	23	ABP29323	Streptococcus poly
621	8	0.6	312	23	AAU93042	Arabidopsis transc
622	8	0.6	316	21	AAG20321	Arabidopsis thalia
623	8	0.6	317	18	AAW23279	Bordetella pertuss
624	8	0.6	324	19	AAW70235	Leishmania antigen
625	8	0.6	324	23	AAB71285	L. chagasi LcgSP10
626	8	0.6	324	23	AAE24949	Leishmania chagasi
627	8	0.6	324	23	ABG60887	Leishmania antigen
628	8	0.6	324	23	AAU71831	Leishmania antigen
629	8	0.6	325	21	AAG20320	Arabidopsis thalia
630	8	0.6	326	21	AAG24814	Arabidopsis thalia
631	8	0.6	326	23	AAU99336	Truncated human ki
632	8	0.6	326	23	ABB08248	Human HsKifC2 moto
633	8	0.6	327	22	ABB70382	Drosophila melanog
634	8	0.6	328	21	AAG46428	Arabidopsis thalia
635	8	0.6	331	21	AAG43812	Arabidopsis thalia
636	8	0.6	332	21	AAG40822	Zea mays protein f
637	8	0.6	337	21	AAG43811	Arabidopsis thalia
638	8	0.6	339	21	AAG54170	Arabidopsis thalia
639	8	0.6	339	22	ABU53124	Intracellular traf
640	8	0.6	343	21	AAG40821	Zea mays protein f
641	8	0.6	349	21	AAG23402	Arabidopsis thalia
642	8	0.6	351	22	ABG08730	Novel human diagno
643	8	0.6	352	22	ABG10202	Novel human diagno
644	8	0.6	354	21	AAG43810	Arabidopsis thalia
645	8	0.6	354	22	AAB30793	Amino acid sequenc
646	8	0.6	356	21	AAG12889	Arabidopsis thalia
647	8	0.6	360	21	AAG12888	Arabidopsis thalia
648	8	0.6	363	21	AAG12887	Arabidopsis thalia
649	8	0.6	363	21	AAG40820	Zea mays protein f
650	8	0.6	367	21	AAG54169	Arabidopsis thalia
651	8	0.6	386	22	ABB70630	Drosophila melanog
652	8	0.6	402	22	ABB70534	Drosophila melanog
653	8	0.6	403	20	AAY28916	Human regulatory p
654	8	0.6	406	18	AAW26523	YC1 polypeptide.
655	8	0.6	406	22	ABB58747	Drosophila melanog
656	8	0.6	408	22	ABB61479	Drosophila melanog
657	8	0.6	408	22	AAE10837	Drosophila melanog
658	8	0.6	413	17	AAR95247	Transcription fact
659	8	0.6	413	19	AAW48390	Homo sapiens E2F4
660	8	0.6	419	23	AAU93147	Arabidopsis transc
661	8	0.6	420	23	ABP73816	Candida albicans e
662	8	0.6	428	24	ABR38876	Surface exposed im
663	8	0.6	428	24	ABP57973	Ferric-siderophore
664	8	0.6	431	21	AAY91277	Group B Streptococ
665	8	0.6	435	22	ABB60210	Drosophila melanog
666	8	0.6	443	14	AAR43895	hEL-nl. Homo sapi

667	8	0.6	443	22	ABG26895	Novel human diagno	740	8	0.6	787	23	ABB08247	Human HsKifC2. Ho
668	8	0.6	444	19	AAW60579	Drosophila neuron-	741	8	0.6	797	20	AAY05850	Banana ripening fr
669	8	0.6	445	21	AAB00060	VRN2 polypeptide.	742	8	0.6	798	22	ABB62356	Drosophila melanog
670	8	0.6	446	19	AAW69563	Sugarbeet choline	743	8	0.6	809	23	ABP73854	Candida albicans e
671	8	0.6	448	22	ABB58601	Drosophila melanog	744	8	0.6	811	22	ABG23352	Novel human diagno
672	8	0.6	461	12	AAR14150	MSP encoded by pUC	745	8	0.6	812	21	AAY95985	Arabidopsis GABA r
673	8	0.6	461	12	AAR14530	Usp45 protein. La	746	8	0.6	814	22	ABB65317	Drosophila melanog
674	8	0.6	464	23	ABB93400	Herbicidally activ	747	8	0.6	825	22	ABB59301	Drosophila melanog
675	8	0.6	464	23	AAM47592	Drosophila cell cy	748	8	0.6	850	22	ABB65764	Drosophila melanog
676	8	0.6	469	22	ABB61392	Drosophila melanog	749	8	0.6	854	22	AAG67391	Partial amino acid
677	8	0.6	478	21	AAB07288	Human IL-2/ P. yoe	750	8	0.6	865	23	ABP35641	Fungal ZBC protein
678	8	0.6	489	22	ABB70861	Drosophila melanog	751	8	0.6	865	23	ABP35642	Fungal ZBC protein
679	8	0.6	497	21	AAG30660	Arabidopsis thalia	752	8	0.6	875	19	AAW34987	Bankia gouldi endo
680	8	0.6	498	21	AAG54168	Arabidopsis thalia	753	8	0.6	880	22	AAB70938	S. pombe potassium
681	8	0.6	510	21	AAG39697	Arabidopsis thalia	754	8	0.6	893	22	ABB64865	Drosophila melanog
682	8	0.6	525	23	ABG91725	Purine/pyrimidine	755	8	0.6	900	22	ABB60688	Drosophila melanog
683	8	0.6	528	23	ABP73734	Candida albicans e	756	8	0.6	916	22	ABB63615	Drosophila melanog
684	8	0.6	529	23	ABG77458	Selected Interacti	757	8	0.6	927	21	AAY95984	Arabidopsis GABA r
685	8	0.6	530	23	ABJ11360	Yeast selected int	758	8	0.6	938	22	ABB60562	Drosophila melanog
686	8	0.6	531	22	ABB70528	Drosophila melanog	759	8	0.6	944	21	AAB43351	Human ORFX ORF3115
687	8	0.6	535	22	ABB58106	Drosophila melanog	760	8	0.6	945	23	ABP73277	Candida albicans e
688	8	0.6	535	22	ABB58915	Drosophila melanog	761	8	0.6	951	22	ABB67950	Drosophila melanog
689	8	0.6	535	23	ABB92665	Herbicidally activ	762	8	0.6	959	22	AAG64950	Human Cream1 prote
690	8	0.6	537	23	AAO14997	Rhodotorula minuta	763	8	0.6	970	23	ABP35631	Fungal ZBC protein
691	8	0.6	540	22	AAG84948	Shrimp white spot	764	8	0.6	976	22	ABB65498	Drosophila melanog
692	8	0.6	541	22	ABB68621	Drosophila melanog	765	8	0.6	984	9	AAP80550	Malaria-specific 1
693	8	0.6	545	22	ABB70472	Drosophila melanog	766	8	0.6	984	13	AAR22247	SERP antigen. Pla
694	8	0.6	554	19	AAW69223	HAP4 protein. Sac	767	8	0.6	984	13	AAR27529	Plasmodium falcipa
695	8	0.6	559	24	ABU02325	S. pneumoniae type	768	8	0.6	984	15	AAR60164	SERA protein encod
696	8	0.6	566	24	ABR40876	Arabidopsis thalia	769	8	0.6	984	16	AAR68837	Plasmodium falcipa
697	8	0.6	576	21	AAG38679	Arabidopsis thalia	770	8	0.6	984	20	AAY29589	Plasmodium falcipa
698	8	0.6	581	23	ABP25491	Streptococcus poly	771	8	0.6	998	22	ABB60423	Drosophila melanog
699	8	0.6	586	11	AAR07274	Reverse transcript	772	8	0.6	1003	22	ABB68762	Drosophila melanog
700	8	0.6	586	12	AAR11920	Reverse transcript	773	8	0.6	1008	22	ABB59260	Drosophila melanog
701	8	0.6	586	16	AAR73935	msDNA-Ec67 reverse	774	8	0.6	1015	23	ABP73325	Candida albicans e
702	8	0.6	586	21	AAY76811	E. coli msDNA-Ec67	775	8	0.6	1028	22	ABB62708	Drosophila melanog
703	8	0.6	586	23	AAM49402	Escherichia coli C	776	8	0.6	1032	21	AAG45673	Arabidopsis thalia
704	8	0.6	594	24	ABP81457	Streptococcus pneu	777	8	0.6	1037	22	ABB70288	Drosophila melanog
705	8	0.6	603	24	ABU11606	Human MDDT polypep	778	8	0.6	1045	21	AAG45672	Arabidopsis thalia
706	8	0.6	605	22	AAE04789	Lycopersicon escul	779	8	0.6	1058	22	AAB68341	A human dishevelle
707	8	0.6	605	22	AAB72308	Neoxanthin cleavag	780	8	0.6	1090	21	AAG45671	Arabidopsis thalia
708	8	0.6	627	20	AAY05853	Banana ripening fr	781	8	0.6	1108	22	ABB63413	Drosophila melanog
709	8	0.6	634	22	ABB71624	Drosophila melanog	782	8	0.6	1132	21	AAG32139	Arabidopsis thalia
710	8	0.6	650	22	AAM93793	Human polypeptide,	783	8	0.6	1152	21	AAG32138	Arabidopsis thalia
711	8	0.6	650	23	ABP68957	Human polypeptide	784	8	0.6	1157	22	ABB62647	Drosophila melanog
712	8	0.6	652	18	AAW18010	Plasmodium falcipa	785	8	0.6	1173	22	ABB62245	Drosophila melanog
713	8	0.6	657	22	ABB57783	Drosophila melanog	786	8	0.6	1182	22	ABB64438	Drosophila melanog
714	8	0.6	663	18	AAW29150	Dual-specific muri	787	8	0.6	1200	22	ABB63533	Drosophila melanog
715	8	0.6	667	22	ABB62659	Drosophila melanog	788	8	0.6	1205	20	AAY33683	P. falciparum 1-de
716	8	0.6	668	22	ABB66214	Drosophila melanog	789	8	0.6	1205	21	AAW90882	P. falciparum 1-de
717	8	0.6	673	22	ABB59674	Drosophila melanog	790	8	0.6	1205	23	AAO21863	Isoprenoid related
718	8	0.6	677	22	ABB62562	Drosophila melanog	791	8	0.6	1238	22	ABB61259	Drosophila melanog
719	8	0.6	677	22	ABB62568	Drosophila melanog	792	8	0.6	1242	20	AAY13461	Amino acid sequenc
720	8	0.6	677	22	ABB65183	Drosophila melanog	793	8	0.6	1242	20	AAW93972	Human IRS-1 and IR
721	8	0.6	678	14	AAR42087	D. melanogaster do	794	8	0.6	1242	22	AAB83921	Amino acid sequenc
722	8	0.6	710	23	ABP73976	Candida albicans e	795	8	0.6	1242	24	ABB99797	Amino acid sequenc
723	8	0.6	713	12	AAR13991	Plasmodium falcipa	796	8	0.6	1242	24	ABG72368	Human Insulin-rece
724	8	0.6	717	22	ABB71958	Drosophila melanog	797	8	0.6	1242	24	ABG72369	Human Insulin-rece
725	8	0.6	718	22	ABB66734	Drosophila melanog	798	8	0.6	1242	24	ABG72370	Human Insulin-rece
726	8	0.6	725	23	AAU93157	Arabidopsis transc	799	8	0.6	1242	24	ABG72371	Human Insulin-rece
727	8	0.6	728	22	ABB64748	Drosophila melanog	800	8	0.6	1242	24	ABG72372	Human Insulin-rece
728	8	0.6	743	22	ABB68417	Drosophila melanog	801	8	0.6	1243	16	AAR67708	Insulin receptor s
729	8	0.6	745	22	ABB71016	Drosophila melanog	802	8	0.6	1247	21	AAB18215	Plasmodium falcipa
730	8	0.6	751	14	AAR33424	G6PD. Plasmodium	803	8	0.6	1278	21	AAB18277	Plasmodium falcipa
731	8	0.6	757	21	AAG41423	Arabidopsis thalia	804	8	0.6	1311	22	ABB61234	Drosophila melanog
732	8	0.6	765	22	ABB70099	Drosophila melanog	805	8	0.6	1316	22	ABB69263	Drosophila melanog
733	8	0.6	769	22	ABB58091	Drosophila melanog	806	8	0.6	1321	17	AAR96994	Mouse IRS-2. Mus
734	8	0.6	770	21	AAY50814	D. melanogaster ac	807	8	0.6	1398	21	AAB18292	Plasmodium falcipa
735	8	0.6	772	21	AAG41422	Arabidopsis thalia	808	8	0.6	1402	23	ABP66040	Bifidobacterium lo
736	8	0.6	780	22	AAM42085	Human polypeptide	809	8	0.6	1406	22	ABB71963	Drosophila melanog
737	8	0.6	784	22	ABB71112	Drosophila melanog	810	8	0.6	1412	22	ABB60943	Drosophila melanog
738	8	0.6	787	22	ABB71302	Drosophila melanog	811	8	0.6	1430	22	AAG85029	Shrimp white spot
739	8	0.6	787	23	AAU99335	Human kinesin moto	812	8	0.6	1475	22	AAM40232	Human polypeptide

813	8	0.6	1486	22	ABB59325	Drosophila melanog		886	7	0.5	14	22	AAE05023	Cryptosporidium pa	
814	8	0.6	1499	22	AAM42018	Human polypeptide		887	7	0.5	14	23	ABG96233	Cysteine-containin	
815	8	0.6	1629	22	ABB63122	Drosophila melanog		888	7	0.5	16	21	AAB08561	Peptide identified	
816	8	0.6	1679	22	ABB60498	Drosophila melanog		889	7	0.5	16	22	AAE05034	Cryptosporidium pa	
817	8	0.6	1679	22	ABB60502	Drosophila melanog		890	7	0.5	17	22	ABG59600	Human liver peptid	
818	8	0.6	1703	22	ABB66223	Drosophila melanog		891	7	0.5	17	22	ABB44227	Peptide #11733 enc	
819	8	0.6	1850	22	AAB72210	Modified chicken v		892	7	0.5	17	22	ABB27103	Protein #9102 enco	
820	8	0.6	1980	22	ABB67589	Drosophila melanog		893	7	0.5	17	22	AAM65259	Human brain expres	
821	8	0.6	2023	22	ABB63487	Drosophila melanog		894	7	0.5	17	22	AAM77964	Human bone marrow	
822	8	0.6	2070	22	ABG03703	Novel human diagno		895	7	0.5	17	22	AAM21858	Peptide #8292 enco	
823	8	0.6	2175	22	ABB65698	Drosophila melanog		896	7	0.5	17	22	AAM38182	Peptide #12219 enc	
824	8	0.6	2274	22	ABB58657	Drosophila melanog		897	7	0.5	17	23	ABB74469	DNA repair protein	
825	8	0.6	2314	22	AAB69136	M. catarrhalis les		898	7	0.5	19	22	ABG51474	Human liver peptid	
826	8	0.6	2391	15	AAR55694	Carbamoyl-phosphat		899	7	0.5	19	22	ABB31409	Peptide #4060 enco	
827	8	0.6	2441	22	ABB62231	Drosophila melanog		900	7	0.5	19	22	ABB36623	Peptide #4129 enco	
828	8	0.6	2500	21	AAB18272	Plasmodium falcipa		901	7	0.5	19	22	ABB21958	Protein #3957 enco	
829	8	0.6	2543	22	ABG21295	Novel human diagno		902	7	0.5	19	22	AAM57384	Human brain expres	
830	8	0.6	3005	22	ABB58102	Drosophila melanog		903	7	0.5	19	22	AAM69777	Human bone marrow	
831	8	0.6	3190	22	AAB84634	Amino acid sequenc		904	7	0.5	19	22	AAM17602	Peptide #4036 enco	
832	8	0.6	3232	22	ABB69873	Drosophila melanog		905	7	0.5	19	22	AAM30122	Peptide #4159 enco	
833	8	0.6	3536	22	ABB65480	Drosophila melanog		906	7	0.5	19	22	AAM05262	Peptide #3944 enco	
834	8	0.6	3726	22	ABB63947	Drosophila melanog		907	7	0.5	19	23	ABG39409	Human peptide enco	
835	8	0.6	4152	22	AAG84992	Shrimp white spot		908	7	0.5	20	24	ABJ38224	Human cytomegalovi	
836	8	0.6	6025	22	AAG84939	Shrimp white spot		909	7	0.5	22	18	AAW45350	Peptide #7 bearing	
837	8	0.6	6077	24	ABP95996	White spot syndrom		910	7	0.5	22	22	ABG56539	Human liver peptid	
838	7	0.5	7	21	AAB15678	Single-chain antib		911	7	0.5	22	22	ABB41096	Peptide #8602 enco	
839	7	0.5	10	22	AAM43208	Mycoplasma genital		912	7	0.5	22	22	ABB25149	Protein #7148 enco	
840	7	0.5	10	22	AAM43210	Mycoplasma genital		913	7	0.5	22	22	AAM61952	Human brain expres	
841	7	0.5	10	22	AAM43212	Mycoplasma genital		914	7	0.5	22	22	AAM74756	Human bone marrow	
842	7	0.5	10	22	AAM43214	Mycoplasma genital		915	7	0.5	22	22	AAM20402	Peptide #6836 enco	
843	7	0.5	10	22	AAM43216	Mycoplasma genital		916	7	0.5	22	22	AAM34872	Peptide #8909 enco	
844	7	0.5	10	22	AAM43218	Mycoplasma genital		917	7	0.5	22	23	ABG44553	Human peptide enco	
845	7	0.5	10	22	AAM43220	Mycoplasma genital		918	7	0.5	23	23	AAE22579	Drosophila MBD-lik	
846	7	0.5	10	22	AAM43222	Mycoplasma genital		919	7	0.5	24	23	ABG69758	Polypeptide identi	
847	7	0.5	10	22	AAM43224	Mycoplasma genital		920	7	0.5	25	22	ABG57335	Human liver peptid	
848	7	0.5	10	22	AAM43226	Mycoplasma genital		921	7	0.5	25	22	ABB41900	Peptide #9406 enco	
849	7	0.5	10	22	AAM43228	Mycoplasma genital		922	7	0.5	25	22	AAM62779	Human brain expres	
850	7	0.5	10	22	AAM43230	Mycoplasma genital		923	7	0.5	25	22	AAM75595	Human bone marrow	
851	7	0.5	10	22	AAM43232	Mycoplasma genital		924	7	0.5	25	22	AAM35702	Peptide #9739 enco	
852	7	0.5	10	22	AAM43234	Mycoplasma genital		925	7	0.5	26	21	AAB09850	Endoglucanase frag	
853	7	0.5	10	22	AAM43236	Mycoplasma genital		926	7	0.5	26	22	ABG56388	Human liver peptid	
854	7	0.5	10	22	AAM43238	Mycoplasma genital		927	7	0.5	26	22	ABB40942	Peptide #8448 enco	
855	7	0.5	10	22	AAM43240	Mycoplasma genital		928	7	0.5	26	22	AAM61804	Human brain expres	
856	7	0.5	10	22	AAM43242	Mycoplasma genital		929	7	0.5	26	22	AAM74603	Human bone marrow	
857	7	0.5	10	22	AAM43244	Mycoplasma genital		930	7	0.5	26	22	AAM34715	Peptide #8752 enco	
858	7	0.5	10	22	AAM43246	Mycoplasma genital		931	7	0.5	30	22	ABG58601	Human liver peptid	
859	7	0.5	10	22	AAM43248	Mycoplasma genital		932	7	0.5	30	22	ABG59523	Human liver peptid	
860	7	0.5	10	22	AAM43250	Mycoplasma genital		933	7	0.5	30	22	ABB43208	Peptide #10714 enc	
861	7	0.5	10	22	AAM43252	Mycoplasma genital		934	7	0.5	30	22	ABB44143	Peptide #11649 enc	
862	7	0.5	10	22	AAM43256	Mycoplasma genital		935	7	0.5	30	22	ABB26297	Protein #8296 enco	
863	7	0.5	10	22	AAM43258	Mycoplasma genital		936	7	0.5	30	22	ABB27027	Protein #9026 enco	
864	7	0.5	10	22	AAG83890	Arabidopsis thalia		937	7	0.5	30	22	AAM64114	Human brain expres	
865	7	0.5	10	22	AAG83892	Arabidopsis thalia		938	7	0.5	30	22	AAM76940	Human bone marrow	
866	7	0.5	10	22	AAG83894	Arabidopsis thalia		939	7	0.5	30	22	AAM77875	Human bone marrow	
867	7	0.5	10	22	AAG83896	Arabidopsis thalia		940	7	0.5	30	22	AAM21033	Peptide #7467 enco	
868	7	0.5	10	22	AAG83898	Arabidopsis thalia		941	7	0.5	30	22	AAM21772	Peptide #8206 enco	
869	7	0.5	10	22	AAG83904	Arabidopsis thalia		942	7	0.5	30	22	AAM37048	Peptide #11085 enc	
870	7	0.5	10	22	AAG83906	Arabidopsis thalia		943	7	0.5	30	22	AAM38096	Peptide #12133 enc	
871	7	0.5	10	22	AAG83908	Arabidopsis thalia		944	7	0.5	30	23	ABG46049	Human peptide enco	
872	7	0.5	10	22	AAG83910	Arabidopsis thalia		945	7	0.5	30	23	ABG46899	Human peptide enco	
873	7	0.5	10	22	AAG83912	Arabidopsis thalia		946	7	0.5	31	22	AAU18263	Novel human DNA-bi	
874	7	0.5	10	22	AAG83914	Arabidopsis thalia		947	7	0.5	31	23	ABG92684	Human DNA-binding	
875	7	0.5	10	22	AAG83923	Arabidopsis thalia		948	7	0.5	32	22	ABG50161	Human liver peptid	
876	7	0.5	10	22	AAG83927	Arabidopsis thalia		949	7	0.5	32	22	ABB30116	Peptide #2767 enco	
877	7	0.5	10	22	AAG84194	Arabidopsis thalia		950	7	0.5	32	22	ABB35288	Peptide #2794 enco	
878	7	0.5	10	22	AAG84198	Arabidopsis thalia		951	7	0.5	32	22	ABB20731	Protein #2730 enco	
879	7	0.5	10	22	AAG84368	Arabidopsis thalia		952	7	0.5	32	22	AAM56117	Human brain expres	
880	7	0.5	10	22	AAG84370	Arabidopsis thalia		953	7	0.5	32	22	AAM68489	Human bone marrow	
881	7	0.5	10	22	AAG87012	Saccharomyces cere		954	7	0.5	32	22	AAM16297	Peptide #2731 enco	
882	7	0.5	10	22	AAG93544	Forkhead related c		955	7	0.5	32	22	AAM28787	Peptide #2824 enco	
883	7	0.5	10	22	AAG93545	Forkhead related c		956	7	0.5	32	22	AAM04032	Peptide #2714 enco	
884	7	0.5	10	22	AAG93546	Forkhead related c		957	7	0.5	32	23	ABG38071	Human peptide enco	
885	7	0.5	13	18	AAW45349	Peptide #6 bearing		958	7	0.5	33	22	ABG50270	Human liver peptid	

959	7	0.5	33	22	ABG58748	Human liver peptid
960	7	0.5	33	22	ABB30242	Peptide #2893 enco
961	7	0.5	33	22	ABB35408	Peptide #2914 enco
962	7	0.5	33	22	ABB20845	Protein #2844 enco
963	7	0.5	33	22	AAM56232	Human brain expres
964	7	0.5	33	22	AAM68610	Human bone marrow
965	7	0.5	33	22	AAM16417	Peptide #2851 enco
966	7	0.5	33	22	AAM28916	Peptide #2953 enco
967	7	0.5	33	22	AAM04151	Peptide #2833 enco
968	7	0.5	33	23	ABG38189	Human peptide enco
969	7	0.5	35	21	AAG11032	Arabidopsis thalia
970	7	0.5	36	21	AAG14995	Arabidopsis thalia
971	7	0.5	40	22	AAO04685	Human polypeptide
972	7	0.5	42	22	AAO04746	Human polypeptide
973	7	0.5	44	21	AAG47925	Arabidopsis thalia
974	7	0.5	44	22	ABG54345	Human liver peptid
975	7	0.5	44	22	ABB39366	Peptide #6872 enco
976	7	0.5	44	22	AAM60044	Human brain expres
977	7	0.5	44	22	AAM72647	Human bone marrow
978	7	0.5	44	22	AAM32881	Peptide #6918 enco
979	7	0.5	44	23	ABG42471	Human peptide enco
980	7	0.5	45	17	AAW08985	Lysine/Serine cont
981	7	0.5	46	21	AAG57114	Arabidopsis thalia
982	7	0.5	46	22	ABG57063	Human liver peptid
983	7	0.5	46	22	ABB69029	Drosophila melanog
984	7	0.5	46	22	ABB41622	Peptide #9128 enco
985	7	0.5	46	22	AAM62493	Human brain expres
986	7	0.5	46	22	AAM75302	Human bone marrow
987	7	0.5	46	22	AAM35415	Peptide #9452 enco
988	7	0.5	48	21	AAG08479	Arabidopsis thalia
989	7	0.5	48	21	AAG47962	Arabidopsis thalia
990	7	0.5	50	23	ABP10282	Human ORFX protein
991	7	0.5	51	22	AAO07319	Human polypeptide
992	7	0.5	52	22	ABG53747	Human liver peptid
993	7	0.5	52	22	ABB38853	Peptide #6359 enco
994	7	0.5	52	22	AAM59498	Human brain expres
995	7	0.5	52	22	AAM72063	Human bone marrow
996	7	0.5	52	22	AAM32328	Peptide #6365 enco
997	7	0.5	52	23	ABG41878	Human peptide enco
998	7	0.5	53	22	ABG11731	Novel human diagno
999	7	0.5	53	22	ABG20393	Novel human diagno
1000	7	0.5	53	22	AAO02332	Human polypeptide

ALIGNMENTS

```
RESULT 1
AAR41989
ID   AAR41989 standard; protein; 190 AA.
XX
AC   AAR41989;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   This sequence represents the Staphylococcus enterotoxin B (SEB)
CC   super antigen (SAg).  Mutant versions of this sequence generated
CC   by random mutation by PCR, may be used to modify the T-cell response
CC   elicited by an antigen, by interacting with specific Vbeta elements
CC   of the T-cell receptors.  This T-cell response modulation may be
CC   useful in a wide variety of autoimmune diseases, where self-reactive
CC   T-cells may be activated by SAg that bind to particular Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 2
AAR41990
ID   AAR41990 standard; protein; 190 AA.
XX
AC   AAR41990;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BR-75.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
```

```
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 3
AAR41991
ID   AAR41991 standard; protein; 190 AA.
XX
AC   AAR41991;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BR-210.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 4
AAR41992
ID   AAR41992 standard; protein; 190 AA.
XX
AC   AAR41992;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BR-257.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;
```

```
  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 5
AAR41993
ID   AAR41993 standard; protein; 190 AA.
XX
AC   AAR41993;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BR-291.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 6
AAR41994
ID   AAR41994 standard; protein; 190 AA.
XX
AC   AAR41994;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BC-6.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 7
AAR41995
ID   AAR41995 standard; protein; 190 AA.
XX
AC   AAR41995;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
```

```
XX
DE   Staphylococcus enterotoxin B mutant BC-66.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 8
AAR41996
ID   AAR41996 standard; protein; 190 AA.
XX
AC   AAR41996;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BC-88.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 9
AAR41997
ID   AAR41997 standard; protein; 190 AA.
XX
AC   AAR41997;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BR-358.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
```

```
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 10
AAR41998
ID   AAR41998 standard; protein; 190 AA.
XX
AC   AAR41998;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BR-374.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 11
AAR41999
ID   AAR41999 standard; protein; 190 AA.
XX
AC   AAR41999;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BA-3.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
```

```
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187

RESULT 12
AAR42000
ID   AAR42000 standard; protein; 190 AA.
XX
AC   AAR42000;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BA-15.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187

RESULT 13
AAR42001
ID   AAR42001 standard; protein; 190 AA.
XX
AC   AAR42001;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BA-24.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              ||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187
```

```
RESULT 14
AAR42002
ID   AAR42002 standard; protein; 190 AA.
XX
AC   AAR42002;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BA-31.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 15
AAR42003
ID   AAR42003 standard; protein; 190 AA.
XX
AC   AAR42003;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BA-50.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 16
AAR42004
ID   AAR42004 standard; protein; 190 AA.
XX
AC   AAR42004;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BA-53.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
```

```
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 17
AAR42005
ID   AAR42005 standard; protein; 190 AA.
XX
AC   AAR42005;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BA-62.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 18
AAR42006
ID   AAR42006 standard; protein; 190 AA.
XX
AC   AAR42006;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BB-14.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
```

```
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 19
AAR42007
ID   AAR42007 standard; protein; 190 AA.
XX
AC   AAR42007;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BB-21.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 20
AAR42008
ID   AAR42008 standard; protein; 190 AA.
XX
AC   AAR42008;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BB-47.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
```

```
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187

RESULT 21
AAR42009
ID   AAR42009 standard; protein; 190 AA.
XX
AC   AAR42009;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BR-30.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187

RESULT 22
AAR42010
ID   AAR42010 standard; protein; 190 AA.
XX
AC   AAR42010;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BR-311.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187

RESULT 23
AAR42011
ID   AAR42011 standard; protein; 190 AA.
XX
AC   AAR42011;
XX
```

```
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BR-474.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 24
AAR42012
ID   AAR42012 standard; protein; 190 AA.
XX
AC   AAR42012;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BA-72.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 25
AAR42013
ID   AAR42013 standard; protein; 190 AA.
XX
AC   AAR42013;
XX
DT   25-MAR-2003  (updated)
DT   21-APR-1994  (first entry)
XX
DE   Staphylococcus enterotoxin B mutant BA-267.
XX
KW   Staphylococcus; enterotoxin B; SEB; super antigen; SAg; receptor;
KW   random mutation; PCR; modify; T-cell; response; antigen; Vbeta element;
KW   autoimmune disease.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   Region          97..157
FT                   /note= "Region undisclosed in the specification"
FT   Region          162..186
FT                   /note= "Region undisclosed in the specification"
XX
PN   WO9314634-A1.
```

```
XX
PD   05-AUG-1993.
XX
PF   28-JAN-1993;   93WO-US00839.
XX
PR   28-JAN-1992;   92US-0827540.
XX
PA   (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
XX
PI   Kappler JW,  Marrack P;
XX
DR   WPI; 1993-320314/40.
XX
PT   Preventing or treating toxic effects of super antigens - by admin. of
PT   new modified or mutated super antigen which induces antibodies but
PT   not T-cell activation
XX
PS   Disclosure; Fig 3; 54pp; English.
XX
CC   The sequences given in AAR41990-2013 are mutant versions of the
CC   Staphylococcus enterotoxin B (SEB) super antigen (SAg).  These mutants
CC   generated by random mutation by PCR, may be used to modify the T-cell
CC   response elicited by an antigen, by interacting with specific Vbeta
CC   elements of the T-cell receptors.  This T-cell response modulation
CC   may be useful in a wide variety of autoimmune diseases, where self-
CC   reactive T-cells may be activated by SAg that bind to particular
CC   Vbeta types.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   190 AA;

  Query Match             1.3%;  Score 17;  DB 14;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-07;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNNI 758
              |||||||||||||||||
Db        171 NNNNNNNNNNNNNNNNI 187


RESULT 26
AAG49727
ID   AAG49727 standard; Protein; 141 AA.
XX
AC   AAG49727;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62937.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
```

```
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             1.2%;  Score 16;  DB 21;  Length 141;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNN 756
              ||||||||||||||||
Db          2 TNNNNNNNNNNNNNNN 17


RESULT 27
AAG49726
ID   AAG49726 standard; Protein; 150 AA.
XX
AC   AAG49726;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 62936.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
```

```
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
```

```
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             1.2%;  Score 16;  DB 21;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNN 756
              ||||||||||||||||
Db         11 TNNNNNNNNNNNNNNN 26


RESULT 28
ABP73877
ID   ABP73877 standard; Protein; 213 AA.
XX
AC   ABP73877;
XX
DT   30-JAN-2003  (first entry)
XX
DE   Candida albicans essential protein SEQ ID NO 7714.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal.
XX
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US49486.
XX
PR   29-DEC-2000; 2000US-259128P.
PR   20-FEB-2001; 2001US-0792024.
PR   22-AUG-2001; 2001US-314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T,  Jiang B,  Boone C,  Bussey H,  Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
DR   N-PSDB; ABZ32427.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele
PT   of a gene and placing other allele of the gene under conditional
PT   expression  -
XX
PS   Claim 44; SEQ ID NO 7714; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of C. albicans cells and for
CC   treating infection by C. albicans. The present sequence is that of an
CC   essential Candida albicans protein used in the method of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office.
XX
SQ   Sequence   213 AA;

  Query Match             1.2%;  Score 16;  DB 23;  Length 213;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNNNNNN 756
              ||||||||||||||||
Db        172 TNNNNNNNNNNNNNNN 187


RESULT 29
AAR24796
ID   AAR24796 standard; Protein; 256 AA.
XX
AC   AAR24796;
XX
DT   25-MAR-2003  (updated)
DT   02-JAN-1992  (first entry)
XX
DE   Sequence of fibronectin-derived peptide which binds GPIIb-IIIa.
XX
KW   Fibronectin; cell attachment; cell adhesion; adhesive glycoprotein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1
FT                   /note= "bonded to NH2 gp.or N-terminal SQ of AAs"
XX
PN   WO9209200-A1.
XX
PD   11-JUN-1992.
XX
PF   03-DEC-1991;   91WO-US09029.
XX
PR   03-DEC-1990;   90US-0620668.
PR   03-JUL-1991;   91US-0725600.
PR   27-NOV-1991;   91US-0803623.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Bowditch R,  Ginsberg MH,  Plow EF;
XX
DR   WPI; 1992-216714/26.
XX
PT   New polypeptide(s) derived from human fibronectin - promote cell
PT   attachment to substrates, inhibit platelet aggregation and
PT   thrombus formation and modulate coagulation and inflammatory
PT   response
XX
PS   Claim 5; Page 97; 112pp; English.
XX
CC   AAQ25592 comprises bps 901-1506 of a human fibronectin (Fn)encoding DNA
```

```
CC   sequence. It encodes AAs 1255-1456. The primary structure of Fn
CC   contains three different types of internal repeats known as homology
CC   Types I,II and III. A site contg. the Arg-Gly-Asp (RGD) sequence in
CC   the 10th Type III repeat of Fn is known to be involved in cell
CC   adhesive events. The Fn peptides of the invention bind to integrin
CC   receptors, particularly GPIIb-IIIa, at a site which is independent
CC   of the RGD sequence. The new binding site is located at least fifty
CC   AAs upstream of the RGD sequence (see AAR24782). The claimed
CC   peptides are 100 AAs or less (see AAR24783-R24795).AAR24796 comprises
CC   a peptide of formula B-X-Z, where X is Fn AA residues 1351-1456,
CC   B is an NH2 gp. or N-terminal sequence of amino acids, and Z is a
CC   COOH gp. or C-terminal sequence of amino acids no more than 150
CC   residues in length.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   256 AA;

  Query Match             1.2%;  Score 16;  DB 13;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        107 NNNNNNNNNNNNNNNN 122


RESULT 30
AAY72761
ID   AAY72761 standard; Protein; 286 AA.
XX
AC   AAY72761;
XX
DT   31-MAY-2001  (first entry)
XX
DE   Dictyostelium discoideum, DdMEK1 protein.
XX
KW   Mitogen-activated protein kinase kinase; MAPKK; DdMEK1 protein;
KW   disease resistance; stress tolerance; signal transduction.
XX
OS   Dictyostelium discoideum.
XX
FH   Key             Location/Qualifiers
FT   Region          121..134
FT                   /note= "Region between subdomains VII and VIII"
FT   Modified-site   127
FT                   /label= Phosphorylation_site
FT   Modified-site   129
FT                   /label= Phosphorylation_site
FT   Modified-site   133
FT                   /label= Phosphorylation_site
XX
PN   EP1078985-A2.
XX
PD   28-FEB-2001.
XX
PF   25-AUG-2000; 2000EP-0307362.
XX
PR   27-AUG-1999;   99US-0384162.
XX
PA   (MIAC ) CANADA MIN AGRIC & AGRI-FOOD CANADA.
XX
PI   Xing T,  Malik K,  Martin-Heller T,  Miki BL;
XX
DR   WPI; 2001-212735/22.
XX
PT   Novel nucleic acids encoding plant-mitogen activated kinase kinase
PT   derivative which has negative charge at core phosphorylation site,
PT   useful for enhancing stress tolerance or increasing disease resistance
PT   in plants -
XX
PS   Claim 11; Fig 1b; 45pp; English.
XX
CC   The present invention relates to a derivative of plant mitogen-activated
CC   protein kinase kinase (MAPKK or MEK) which contains a negative charge at
CC   the core phosphorylation site of MAPKK. MAPKK activates MAPK signalling
CC   cascades which are ubiquitous among eukaryotes from yeast to human and
CC   mediate a large array of signal transduction pathways in plants.
CC   Mutagenesis of a core phosphorylation site of a member of MAPK cascade
CC   create permanently active forms, which stimulate both pathogen- and
CC   wound- inducible genes when introduced into plant cells. The MAPK gene is
CC   useful for increasing disease resistance or enhancing stress tolerance
CC   in a plant, preferably tomato cv Bonny Best. The MEK gene is introduced
CC   into the plant by transformation or particle bombardment.
CC   The present sequence is Dictyostelium discoideum, DdMEK1 protein.
CC   This sequence is a member of mitogen-activated protein kinase kinase
CC   (MAPKK) family.
CC   Note: This sequence designated as SEQ ID NO: 6 is stated as being the
CC   same as that shown in sequence listing (AAY72779) of the specification.
CC   However, the sequences differ at several positions.
XX
SQ   Sequence   286 AA;

  Query Match             1.2%;  Score 16;  DB 22;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        206 NNNNNNNNNNNNNNNN 221


RESULT 31
AAW03625
ID   AAW03625 standard; peptide; 342 AA.
XX
AC   AAW03625;
XX
DT   25-MAR-2003  (updated)
DT   17-DEC-1996  (first entry)
XX
DE   Human lutropin-choriogonadotropin GPR N-terminal sequence.
XX
KW   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
KW   muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
KW   odorant; cytomegalovirus; serotonergic.
XX
OS   Homo sapiens.
XX
PN   US5508384-A.
XX
PD   16-APR-1996.
XX
PF   09-SEP-1993;   93US-0118270.
XX
PR   09-SEP-1993;   93US-0118270.
PR   10-SEP-1992;   92US-0943236.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
XX
PI   Murphy RB,  Schuster DI;
XX
DR   WPI; 1996-208785/21.
XX
PT   New dopamine receptor peptide - useful as antipsychotic agent, e.g.
PT   for treating schizophrenia
XX
PS   Disclosure; Fig 8B(2); 184pp; English.
XX
CC   Proteins AAW02657-W02720 represent a range of G-protein coupled receptor
CC   (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other GPR proteins.  The peptides
CC   AAW03578-W03651 represent the N-terminal fragments of the above
CC   proteins. The receptor proteins were used to design polypeptides, pref.
CC   based on the transmembrane domains, for use in G-protein coupled receptor
```

```
CC   ligand binding assays.  The polypeptide fragments retain biological
CC   activity such as binding a GPR ligand or modulating GPR ligand binding to
CC   a GPR (see AAW02747-W02999 for examples of polypeptide fragments).
CC   The polypeptide fragments can be used in compositions for treating
CC   subjects suffering from a pathology related to a GPR abnormality e.g. a
CC   psychotic disorder such as schizophrenia.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   342 AA;

  Query Match             1.2%;  Score 16;  DB 17;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db         34 NNNNNNNNNNNNNNNN 49


RESULT 32
AAG16708
ID   AAG16708 standard; Protein; 346 AA.
XX
AC   AAG16708;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 17457.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
```

```
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             1.2%;  Score 16;  DB 21;  Length 346;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        230 NNNNNNNNNNNNNNNN 245


RESULT 33
AAY82392
ID   AAY82392 standard; Protein; 349 AA.
XX
AC   AAY82392;
XX
DT   27-JUN-2000  (first entry)
XX
DE   C. trachomatis MOMP containing fusion protein SEQ ID NO:5.
XX
KW   Chlamydia trachomatis; fusion protein; major outer membrane protein;
KW   MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;
KW   infection; infectious disease.
XX
OS   Chlamydia trachomatis.
OS   Synthetic.
XX
PN   JP2000041678-A.
XX
PD   15-FEB-2000.
XX
PF   28-JUL-1998;   98JP-0213212.
XX
PR   28-JUL-1998;   98JP-0213212.
XX
PA   (ELED ) DENKI KAGAKU KOGYO KK.
XX
DR   WPI; 2000-295780/26.
DR   N-PSDB; AAA08124.
XX
PT   A soluble fused protein useful for diagnosis of Chlamydia infection,
PT   comprises at least part of major outer membrane protein (MOMP) of
PT   Chlamydia trachomatis -
XX
PS   Claim 9; Page 24-25; 37pp; Japanese.
XX
CC   The present invention describes fusion proteins (I) comprising at least
CC   part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,
CC   at least one hydrophilic polypeptide having no immunoreactivity to
CC   human serum and their connected part. AAA08120 to AAA08125 encode
CC   specifically claimed examples of the fusion proteins given in AAY82388
CC   to AAY82393. Also described is a method (A) for the detection of
CC   Chlamydia trachomatis antibody using (I) as the antigen. (I) is used for
CC   the diagnosis of Chlamydia trachomatis infectious diseases. The method
CC   can diagnose Chlamydia trachomatis infectious diseases specifically in a
CC   high sensitivity.
XX
SQ   Sequence   349 AA;

  Query Match             1.2%;  Score 16;  DB 21;  Length 349;
  Best Local Similarity   100.0%;  Pred. No. 2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        168 NNNNNNNNNNNNNNNN 183


RESULT 34
AAO19019
ID   AAO19019 standard; Protein; 357 AA.
```

```
XX
AC   AAO19019;
XX
DT   16-DEC-2002  (first entry)
XX
DE   Human kinesin motor protein HsKif16a motor domain.
XX
KW   Human; kinesin motor protein; kinesin; HsKif16a; cancer; motor domain;
KW   vesicular transport disorder; neurological disorder; immune disorder;
KW   inflammation; cytostatic; neuroprotective; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 16
FT                   /note= "encoded by AGN"
FT   Misc-difference 17
FT                   /note= "encoded by NNN"
FT   Misc-difference 18
FT                   /note= "encoded by NNN"
FT   Misc-difference 19
FT                   /note= "encoded by NNN"
FT   Misc-difference 20
FT                   /note= "encoded by NNN"
FT   Misc-difference 21
FT                   /note= "encoded by NNN"
FT   Misc-difference 22
FT                   /note= "encoded by NNN"
FT   Misc-difference 23
FT                   /note= "encoded by NNN"
FT   Misc-difference 24
FT                   /note= "encoded by NNN"
FT   Misc-difference 25
FT                   /note= "encoded by NNN"
FT   Misc-difference 26
FT                   /note= "encoded by NNN"
FT   Misc-difference 27
FT                   /note= "encoded by NNN"
FT   Misc-difference 28
FT                   /note= "encoded by NNN"
FT   Misc-difference 29
FT                   /note= "encoded by NNN"
FT   Misc-difference 30
FT                   /note= "encoded by NNN"
FT   Misc-difference 31
FT                   /note= "encoded by NNN"
FT   Misc-difference 32
FT                   /note= "encoded by NNN"
FT   Misc-difference 33
FT                   /note= "encoded by NNN"
FT   Misc-difference 34
FT                   /note= "encoded by NNN"
FT   Misc-difference 35
FT                   /note= "encoded by NNN"
FT   Misc-difference 36
FT                   /note= "encoded by NNN"
FT   Misc-difference 37
FT                   /note= "encoded by NNN"
XX
PN   US6420162-B1.
XX
PD   16-JUL-2002.
XX
PF   22-NOV-2000; 2000US-0718810.
XX
PR   22-NOV-2000; 2000US-0718810.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Beraud C,  Freedman R;
XX
DR   WPI; 2002-664560/71.
DR   N-PSDB; AAL49367.
XX
PT   New nucleic acid encoding a motor protein which has microtubule
PT   stimulated ATPase activity, designated HSKIF16A, for preventing and
PT   treating cancer, and neurological and vesicle transport disorders  -
XX
PS   Claim 1; Column 35-36; 27pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   human motor protein which has microtubule stimulated ATPase activity and
CC   is designated HSKIF16A. The nucleic acid is used to diagnose, treat and
CC   prevent cancer, immune disorders, inflammation, neurological disorders
CC   and disorders of vesicular transport. The present sequence is the motor
CC   domain of the protein of the invention.
XX
SQ   Sequence   357 AA;

  Query Match              1.2%;  Score 16;  DB 23;  Length 357;
  Best Local Similarity   100.0%;  Pred. No. 2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTMLG 107
              ||||||||||||||||
Db         98 FAYGQTGSGKTYTMLG 113


RESULT 35
AAU74558
ID   AAU74558 standard; Protein; 357 AA.
XX
AC   AAU74558;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human kinesin motor protein HsKif16a motor domain.
XX
KW   Human; kinesin motor protein; HsKif16a; motor domain; inflammation;
KW   microtubule-stimulated ATPase activity; adenosine triphosphatase; cancer;
KW   neurological disorder; vesicular transport disorder; autoimmune disease;
KW   arthritis; graft rejection; proliferation; cytostatic; neurotropic;
KW   immunosuppressive; antiarthritic; antiinflammatory.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 16
FT                   /label= unknown
FT                   /note= "Encoded by AGN"
FT   Misc-difference 17..37
FT                   /label= unknown
FT                   /note= "Encoded by 21(NNN)"
FT   Misc-difference 357
FT                   /note= "Encoded by AT"
XX
PN   US6333184-B1.
XX
PD   25-DEC-2001.
XX
PF   22-NOV-2000; 2000US-0718841.
XX
PR   22-NOV-2000; 2000US-0718841.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Beraud C,  Freedman R;
XX
DR   WPI; 2002-163180/21.
DR   N-PSDB; ABK14094.
XX
PT   New human kinesin motor protein, HsKif16a, having
PT   microtubule-stimulated ATPase activity, useful for identifying specific
PT   modulators for e.g. treating cancer  -
XX
PS   Claim 3; Fig 4; 27pp; English.
```

```
XX
CC   The invention relates to a human kinesin motor protein, HsKif16a,
CC   comprising a motor domain and having microtubule-stimulated ATPase
CC   (adenosine triphosphatase) activity. The protein of the invention and its
CC   associated nucleic acid are useful in diagnosis, treatment and prevention
CC   of cancer, neurological disorders, disorders of vesicular transport,
CC   autoimmune diseases, arthritis, graft rejection, inflammation and
CC   proliferation induced after medical procedures. The protein is also used
CC   to identify its specific modulators, potentially useful as therapeutic
CC   agents. This sequence represents the motor domain of the HsKif16a
CC   polypeptide of the invention.
XX
SQ   Sequence   357 AA;

  Query Match              1.2%;  Score 16;  DB 23;  Length 357;
  Best Local Similarity   100.0%;  Pred. No. 2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTMLG 107
              ||||||||||||||||
Db         98 FAYGQTGSGKTYTMLG 113


RESULT 36
AAG16707
ID   AAG16707 standard; Protein; 376 AA.
XX
AC   AAG16707;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 17456.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
```

```
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             1.2%;  Score 16;  DB 21;  Length 376;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        260 NNNNNNNNNNNNNNNN 275


RESULT 37
AAG16706
ID   AAG16706 standard; Protein; 378 AA.
XX
AC   AAG16706;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 17455.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
```

```
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.
```

```
  Query Match             1.2%;  Score 16;  DB 21;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        262 NNNNNNNNNNNNNNNN 277


RESULT 38
AAB18414
ID   AAB18414 standard; Protein; 432 AA.
XX
AC   AAB18414;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Amino acid sequence of a CelF cellulase protein.
XX
KW   CelF; cellulase; endoglucanase; cellobiose; glucose; cellobiose;
KW   cellulosic material; textile; paper; ethanol; fuel;
KW   cellulosic feed stock.
XX
OS   Orpinomyces sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "signal peptide"
FT   Protein         22..432
FT                   /note= "mature protein"
FT   Domain          22..57
FT                   /note= "fungal cellulose binding domain"
FT   Peptide         67..105
FT                   /note= "Asn-rich linker sequence"
FT   Domain          106..432
FT                   /note= "catalytic domain"
XX
PN   US6114158-A.
XX
PD   05-SEP-2000.
XX
PF   17-JUL-1998;   98US-0118319.
XX
PR   17-JUL-1998;   98US-0118319.
XX
PA   (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Li X,  Ljungdahl LG,  Chen H;
XX
DR   WPI; 2000-593548/56.
DR   N-PSDB; AAA75403.
XX
PT   Recombinant DNA molecule encoding a mature CelF cellulase, useful for
PT   degrading cellulosic material in textile and paper industries, and for
PT   ethanol production  -
XX
PS   Disclosure; Column 13-16; 26pp; English.
XX
CC   The present sequence represents a CelF protein of Orpinomyces sp. PC-2.
CC   CelF is a cellulase, which has endoglucanase activity and produces
CC   cellobiose from cellotetroase or cellopentoase, and glucose and
CC   cellobiose from cellotriose. The CelF polynucleotide is useful for
CC   recombinantly expressing CelF mature protein in Escherichia coli or
CC   other host cells. The cellulase protein is useful for degrading
CC   cellulosic material in textile and paper industries, and also for
CC   producing ethanol. Cellulase is also useful for producing fuels and
CC   chemicals from cellulosic feed stocks.
XX
SQ   Sequence   432 AA;

  Query Match             1.2%;  Score 16;  DB 21;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db         78 NNNNNNNNNNNNNNNN 93


RESULT 39
AAB30815
ID   AAB30815 standard; Protein; 477 AA.
XX
AC   AAB30815;
XX
DT   02-APR-2001  (first entry)
XX
DE   Amino acid sequence of a prion-like amyloidogenic protein.
XX
KW   SCHAG; self-coalesce; higher-order aggregate; amyloidogenic domain;
KW   aggregation; fibril; phenotypic alteration; gene therapy;
KW   disease resistance; plant pigmentation; prion disease.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200075324-A2.
XX
PD   14-DEC-2000.
XX
PF   09-JUN-2000; 2000WO-US15876.
XX
PR   09-JUN-1999;   99US-0138833.
XX
PA   (ARCH-) ARCH DEV CORP.
XX
PI   Lindquist S,  Li L,  Ma J,  Liu J,  Sondheimer N,  Scheibel T;
XX
DR   WPI; 2001-061723/07.
XX
PT   New nucleic acid encoding chimeric proteins with self-assembly
PT   properties, useful e.g. for diagnosis and treatment of prion diseases,
PT   also related aggregates, fibrils and polymers -
XX
PS   Claim 11; Page 162-163; 188pp; English.
XX
CC   AAB30803-17 represent prion-like amyloidogenic proteins which have
CC   been identified from a yeast protein database. The specification
CC   describes chimeric polypeptides, which comprise at least one SCHAG
CC   (self-coalesces into higher-order aggregates) amino acid sequence fused
CC   in frame with a polypeptide of interest (which is other than a marker
CC   protein, a glutathione-S-transferase or a staphylococcal nuclear
CC   protein). The specification also describes chimeric polypeptides that
CC   comprises an amyloidogenic domain that causes aggregation into fibrils.
CC   The chimeric polypeptides are used to prepare polymers with multiple
CC   reactivities, e.g. derivatised with enzymes, or specific binding
CC   partners, and useful e.g. for performing multi-step chemical reactions.
CC   They can be used create an inducible, or stable phenotypic alteration
CC   in a cell, e.g. for gene therapy, protein production, imparting disease
CC   resistance to plants, altering plant pigmentation and for diagnosis
CC   and treatment of prion diseases.
XX
SQ   Sequence   477 AA;

  Query Match             1.2%;  Score 16;  DB 22;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        244 NNNNNNNNNNNNNNNN 259


RESULT 40
AAO19018
ID   AAO19018 standard; Protein; 563 AA.
XX
```

```
AC   AAO19018;
XX
DT   16-DEC-2002  (first entry)
XX
DE   Human kinesin motor protein HsKif16a.
XX
KW   Human; kinesin motor protein; kinesin; HsKif16a; cancer;
KW   vesicular transport disorder; neurological disorder; immune disorder;
KW   inflammation; cytostatic; neuroprotective; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 16
FT                   /note= "encoded by AGN"
FT   Misc-difference 17
FT                   /note= "encoded by NNN"
FT   Misc-difference 18
FT                   /note= "encoded by NNN"
FT   Misc-difference 19
FT                   /note= "encoded by NNN"
FT   Misc-difference 20
FT                   /note= "encoded by NNN"
FT   Misc-difference 21
FT                   /note= "encoded by NNN"
FT   Misc-difference 22
FT                   /note= "encoded by NNN"
FT   Misc-difference 23
FT                   /note= "encoded by NNN"
FT   Misc-difference 24
FT                   /note= "encoded by NNN"
FT   Misc-difference 25
FT                   /note= "encoded by NNN"
FT   Misc-difference 26
FT                   /note= "encoded by NNN"
FT   Misc-difference 27
FT                   /note= "encoded by NNN"
FT   Misc-difference 28
FT                   /note= "encoded by NNN"
FT   Misc-difference 29
FT                   /note= "encoded by NNN"
FT   Misc-difference 30
FT                   /note= "encoded by NNN"
FT   Misc-difference 31
FT                   /note= "encoded by NNN"
FT   Misc-difference 32
FT                   /note= "encoded by NNN"
FT   Misc-difference 33
FT                   /note= "encoded by NNN"
FT   Misc-difference 34
FT                   /note= "encoded by NNN"
FT   Misc-difference 35
FT                   /note= "encoded by NNN"
FT   Misc-difference 36
FT                   /note= "encoded by NNN"
FT   Misc-difference 37
FT                   /note= "encoded by NNN"
FT   Misc-difference 391
FT                   /note= "encoded by NNN"
FT   Misc-difference 392
FT                   /note= "encoded by NNN"
FT   Misc-difference 393
FT                   /note= "encoded by NNN"
FT   Misc-difference 394
FT                   /note= "encoded by NNN"
FT   Misc-difference 395
FT                   /note= "encoded by NNN"
FT   Misc-difference 396
FT                   /note= "encoded by NNN"
FT   Misc-difference 397
FT                   /note= "encoded by NNN"
FT   Misc-difference 398
FT                   /note= "encoded by NNN"
FT   Misc-difference 399
FT                   /note= "encoded by NNN"
FT   Misc-difference 400
FT                   /note= "encoded by NNN"
FT   Misc-difference 401
FT                   /note= "encoded by NNN"
FT   Misc-difference 402
FT                   /note= "encoded by NNN"
FT   Misc-difference 403
FT                   /note= "encoded by NNN"
FT   Misc-difference 404
FT                   /note= "encoded by NNN"
FT   Misc-difference 405
FT                   /note= "encoded by NNN"
FT   Misc-difference 406
FT                   /note= "encoded by NNN"
FT   Misc-difference 407
FT                   /note= "encoded by NNN"
FT   Misc-difference 408
FT                   /note= "encoded by NNN"
FT   Misc-difference 409
FT                   /note= "encoded by NNN"
FT   Misc-difference 410
FT                   /note= "encoded by NNN"
FT   Misc-difference 411
FT                   /note= "encoded by NNN"
FT   Misc-difference 412
FT                   /note= "encoded by NNN"
XX
PN   US6420162-B1.
XX
PD   16-JUL-2002.
XX
PF   22-NOV-2000; 2000US-0718810.
XX
PR   22-NOV-2000; 2000US-0718810.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Beraud C,  Freedman R;
XX
DR   WPI; 2002-664560/71.
DR   N-PSDB; AAL49366.
XX
PT   New nucleic acid encoding a motor protein which has microtubule
PT   stimulated ATPase activity, designated HSKIF16A, for preventing and
PT   treating cancer, and neurological and vesicle transport disorders  -
XX
PS   Claim 1; Column 29-34; 27pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   human motor protein which has microtubule stimulated ATPase activity and
CC   is designated HSKIF16A. The nucleic acid is used to diagnose, treat and
CC   prevent cancer, immune disorders, inflammation, neurological disorders
CC   and disorders of vesicular transport. The present sequence is the protein
CC   of the invention.
XX
SQ   Sequence   563 AA;

  Query Match             1.2%;  Score 16;  DB 23;  Length 563;
  Best Local Similarity   100.0%;  Pred. No. 3e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTMLG 107
              ||||||||||||||||
Db         98 FAYGQTGSGKTYTMLG 113


Search completed: October  2, 2003, 16:32:15
Job time : 115 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 2, 2003, 16:30:31 ; Search time 23 Seconds (without alignments) 2369.408 Million cell updates/sec

Title: US-10-006-780-2
Perfect score: 1288
Sequence: 1 MNSKIKVVVRKRPLSELEKK..........KKLVQDNIKNSMDHNNIHKK 1288

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	16	1.2	286	4	US-09-384-162-6	Sequence 6, Appli
2	16	1.2	357	4	US-09-718-841-4	Sequence 4, Appli
3	16	1.2	357	4	US-09-718-810-4	Sequence 4, Appli
4	16	1.2	428	3	US-09-118-319-5	Sequence 5, Appli
5	16	1.2	432	3	US-09-118-319-2	Sequence 2, Appli
6	16	1.2	563	4	US-09-718-841-2	Sequence 2, Appli
7	16	1.2	563	4	US-09-718-810-2	Sequence 2, Appli
8	16	1.2	732	4	US-08-914-999-8	Sequence 8, Appli
9	16	1.2	1311	1	US-08-340-011-5	Sequence 5, Appli
10	16	1.2	1311	3	US-08-901-710-5	Sequence 5, Appli
11	16	1.2	1584	4	US-09-457-040B-27	Sequence 27, Appl
12	16	1.2	2184	4	US-09-417-485D-6	Sequence 6, Appli
13	15	1.2	303	4	US-09-722-129-4	Sequence 4, Appli
14	15	1.2	492	4	US-09-722-129-2	Sequence 2, Appli
15	15	1.2	794	4	US-09-417-485D-8	Sequence 8, Appli
16	15	1.2	888	2	US-08-861-464-6	Sequence 6, Appli
17	15	1.2	888	2	US-08-396-001-6	Sequence 6, Appli
18	15	1.2	888	3	US-09-323-433A-6	Sequence 6, Appli
19	14	1.1	23	3	US-09-098-901-11	Sequence 11, Appl
20	14	1.1	337	4	US-09-641-806-4	Sequence 4, Appli
21	14	1.1	337	4	US-09-723-129-4	Sequence 4, Appli
22	14	1.1	337	4	US-09-722-862-4	Sequence 4, Appli
23	14	1.1	342	4	US-09-641-806-2	Sequence 2, Appli
24	14	1.1	342	4	US-09-723-129-2	Sequence 2, Appli
25	14	1.1	342	4	US-09-722-862-2	Sequence 2, Appli
26	14	1.1	730	4	US-09-291-170A-3	Sequence 3, Appli
27	14	1.1	730	4	US-09-724-884-3	Sequence 3, Appli
28	14	1.1	955	1	US-08-006-676B-1	Sequence 1, Appli
29	14	1.1	955	1	US-08-282-845-2	Sequence 2, Appli
30	14	1.1	955	2	US-08-428-414A-3	Sequence 3, Appli
31	14	1.1	955	5	PCT-US94-00324-1	Sequence 1, Appli
32	13	1.0	341	4	US-09-724-517-4	Sequence 4, Appli
33	13	1.0	341	4	US-09-641-807A-4	Sequence 4, Appli
34	13	1.0	341	4	US-09-723-096-4	Sequence 4, Appli
35	13	1.0	1151	4	US-09-177-165A-31	Sequence 31, Appl
36	13	1.0	1279	4	US-09-724-517-2	Sequence 2, Appli
37	13	1.0	1279	4	US-09-641-807A-2	Sequence 2, Appli
38	13	1.0	1279	4	US-09-723-096-2	Sequence 2, Appli
39	12	0.9	246	3	US-09-451-117-2	Sequence 2, Appli
40	12	0.9	246	4	US-09-888-655-2	Sequence 2, Appli
41	12	0.9	362	4	US-09-594-669-6	Sequence 6, Appli
42	12	0.9	391	4	US-09-594-669-4	Sequence 4, Appli
43	12	0.9	405	4	US-09-594-669-12	Sequence 12, Appl
44	12	0.9	430	4	US-09-594-669-2	Sequence 2, Appli
45	12	0.9	434	4	US-09-594-669-10	Sequence 10, Appl
46	12	0.9	473	4	US-09-594-669-8	Sequence 8, Appli
47	12	0.9	675	1	US-08-386-495-10	Sequence 10, Appl
48	12	0.9	675	5	PCT-US96-02331-10	Sequence 10, Appl
49	12	0.9	678	4	US-09-595-684B-25	Sequence 25, Appl
50	12	0.9	723	4	US-09-594-669-14	Sequence 14, Appl
51	12	0.9	725	4	US-09-594-669-16	Sequence 16, Appl
52	12	0.9	725	4	US-09-595-684B-33	Sequence 33, Appl
53	12	0.9	842	5	PCT-US96-02331-15	Sequence 15, Appl
54	11	0.9	205	4	US-09-637-481-2	Sequence 2, Appli
55	11	0.9	205	4	US-09-723-428-2	Sequence 2, Appli
56	11	0.9	205	4	US-09-724-520-2	Sequence 2, Appli
57	11	0.9	319	4	US-09-724-510-2	Sequence 2, Appli
58	11	0.9	319	4	US-09-723-216-2	Sequence 2, Appli
59	11	0.9	319	4	US-09-675-227-2	Sequence 2, Appli
60	11	0.9	343	4	US-09-967-908A-8	Sequence 8, Appli
61	11	0.9	361	4	US-09-596-541-2	Sequence 2, Appli
62	11	0.9	361	4	US-09-723-595-2	Sequence 2, Appli
63	11	0.9	363	4	US-09-967-908A-10	Sequence 10, Appl
64	11	0.9	369	4	US-09-596-541-6	Sequence 6, Appli
65	11	0.9	369	4	US-09-723-595-6	Sequence 6, Appli
66	11	0.9	370	4	US-09-596-541-4	Sequence 4, Appli
67	11	0.9	370	4	US-09-723-595-4	Sequence 4, Appli
68	11	0.9	373	4	US-09-967-908A-4	Sequence 4, Appli
69	11	0.9	375	4	US-09-572-191-4	Sequence 4, Appli
70	11	0.9	375	4	US-09-723-262-4	Sequence 4, Appli
71	11	0.9	375	4	US-09-723-219-4	Sequence 4, Appli
72	11	0.9	381	4	US-09-967-908A-6	Sequence 6, Appli
73	11	0.9	409	4	US-09-572-191-6	Sequence 6, Appli
74	11	0.9	409	4	US-09-723-262-6	Sequence 6, Appli
75	11	0.9	409	4	US-09-723-219-6	Sequence 6, Appli
76	11	0.9	473	4	US-09-592-054-6	Sequence 6, Appli
77	11	0.9	519	4	US-09-595-684B-37	Sequence 37, Appl
78	11	0.9	522	4	US-09-592-054-4	Sequence 4, Appli
79	11	0.9	935	4	US-09-914-259-25	Sequence 25, Appl
80	11	0.9	1231	4	US-09-595-684B-23	Sequence 23, Appl
81	11	0.9	1232	4	US-09-592-054-2	Sequence 2, Appli
82	11	0.9	1234	4	US-09-592-054-8	Sequence 8, Appli
83	11	0.9	1368	4	US-09-967-908A-2	Sequence 2, Appli
84	11	0.9	1388	4	US-09-572-191-2	Sequence 2, Appli
85	11	0.9	1388	4	US-09-723-262-2	Sequence 2, Appli
86	11	0.9	1388	4	US-09-723-219-2	Sequence 2, Appli
87	10	0.8	24	3	US-09-098-901-9	Sequence 9, Appli
88	10	0.8	40	1	US-08-478-312-54	Sequence 54, Appl
89	10	0.8	40	1	US-08-485-302-54	Sequence 54, Appl
90	10	0.8	40	1	US-08-476-169-59	Sequence 59, Appl
91	10	0.8	40	1	US-08-484-083-59	Sequence 59, Appl
92	10	0.8	40	2	US-08-764-640-238	Sequence 238, App
93	10	0.8	40	3	US-08-973-225-226	Sequence 226, App
94	10	0.8	40	3	US-09-244-298A-238	Sequence 238, App
95	10	0.8	40	3	US-09-516-704-238	Sequence 238, App
96	10	0.8	40	4	US-09-549-090-226	Sequence 226, App
97	10	0.8	40	4	US-09-832-230A-238	Sequence 238, App
98	10	0.8	100	2	US-08-710-249-16	Sequence 16, Appl
99	10	0.8	100	4	US-09-220-157A-16	Sequence 16, Appl
100	10	0.8	274	4	US-09-723-430-2	Sequence 2, Appli

101	10	0.8	274	4	US-09-632-155-2	Sequence 2, Appli
102	10	0.8	274	4	US-09-724-518-2	Sequence 2, Appli
103	10	0.8	359	4	US-09-722-139-4	Sequence 4, Appli
104	10	0.8	359	4	US-09-721-832-4	Sequence 4, Appli
105	10	0.8	359	4	US-09-721-689-4	Sequence 4, Appli
106	10	0.8	412	4	US-09-471-396-1	Sequence 1, Appli
107	10	0.8	412	4	US-09-788-345-12	Sequence 12, Appl
108	10	0.8	522	1	US-08-164-614A-10	Sequence 10, Appl
109	10	0.8	522	2	US-08-456-489B-10	Sequence 10, Appl
110	10	0.8	536	1	US-08-164-614A-12	Sequence 12, Appl
111	10	0.8	536	2	US-08-456-489B-12	Sequence 12, Appl
112	10	0.8	596	4	US-09-752-165-2	Sequence 2, Appli
113	10	0.8	659	2	US-08-709-874A-20	Sequence 20, Appl
114	10	0.8	659	3	US-09-104-382-20	Sequence 20, Appl
115	10	0.8	675	2	US-08-971-036-2	Sequence 2, Appli
116	10	0.8	675	3	US-09-096-570-2	Sequence 2, Appli
117	10	0.8	675	4	US-09-265-617B-2	Sequence 2, Appli
118	10	0.8	690	3	US-08-816-977-47	Sequence 47, Appl
119	10	0.8	694	3	US-08-816-977-49	Sequence 49, Appl
120	10	0.8	708	3	US-08-816-977-33	Sequence 33, Appl
121	10	0.8	711	3	US-08-816-977-35	Sequence 35, Appl
122	10	0.8	809	4	US-09-660-552-5	Sequence 5, Appli
123	10	0.8	1084	1	US-08-717-515-6	Sequence 6, Appli
124	10	0.8	1276	1	US-08-717-515-8	Sequence 8, Appli
125	10	0.8	1375	4	US-09-722-139-2	Sequence 2, Appli
126	10	0.8	1375	4	US-09-721-832-2	Sequence 2, Appli
127	10	0.8	1375	4	US-09-721-689-2	Sequence 2, Appli
128	9	0.7	11	1	US-08-040-548-43	Sequence 43, Appl
129	9	0.7	11	1	US-08-466-344-43	Sequence 43, Appl
130	9	0.7	11	6	5206152-12	Patent No. 5206152
131	9	0.7	23	3	US-09-098-901-7	Sequence 7, Appli
132	9	0.7	177	4	US-09-247-155-164	Sequence 164, App
133	9	0.7	330	4	US-09-134-001C-4002	Sequence 4002, Ap
134	9	0.7	355	4	US-09-724-511-4	Sequence 4, Appli
135	9	0.7	355	4	US-09-723-097-4	Sequence 4, Appli
136	9	0.7	355	4	US-09-632-344-4	Sequence 4, Appli
137	9	0.7	367	4	US-09-724-511-2	Sequence 2, Appli
138	9	0.7	367	4	US-09-723-097-2	Sequence 2, Appli
139	9	0.7	367	4	US-09-632-344-2	Sequence 2, Appli
140	9	0.7	416	4	US-09-457-040B-10	Sequence 10, Appl
141	9	0.7	677	4	US-09-328-352-4365	Sequence 4365, Ap
142	9	0.7	1003	1	US-08-571-758-4	Sequence 4, Appli
143	9	0.7	1003	1	US-08-909-984A-4	Sequence 4, Appli
144	9	0.7	1003	1	US-08-909-983-4	Sequence 4, Appli
145	9	0.7	2938	5	PCT-US94-00198-3	Sequence 3, Appli
146	8	0.6	90	4	US-09-186-276B-59	Sequence 59, Appl
147	8	0.6	90	4	US-08-842-445-59	Sequence 59, Appl
148	8	0.6	90	4	US-09-186-188B-59	Sequence 59, Appl
149	8	0.6	193	2	US-08-679-765-3	Sequence 3, Appli
150	8	0.6	193	2	US-09-196-525-3	Sequence 3, Appli
151	8	0.6	193	3	US-09-318-317-3	Sequence 3, Appli
152	8	0.6	261	6	5320958-5	Patent No. 5320958
153	8	0.6	265	4	US-09-611-659A-4	Sequence 4, Appli
154	8	0.6	272	2	US-08-492-027A-4	Sequence 4, Appli
155	8	0.6	299	3	US-09-621-233-2	Sequence 2, Appli
156	8	0.6	299	4	US-09-724-508-2	Sequence 2, Appli
157	8	0.6	299	4	US-09-724-516-2	Sequence 2, Appli
158	8	0.6	301	4	US-09-134-001C-2915	Sequence 2915, Ap
159	8	0.6	317	1	US-08-221-750A-13	Sequence 13, Appl
160	8	0.6	324	4	US-09-183-861-53	Sequence 53, Appl
161	8	0.6	324	4	US-09-022-765-53	Sequence 53, Appl
162	8	0.6	324	4	US-09-551-974A-53	Sequence 53, Appl
163	8	0.6	326	4	US-09-721-383-4	Sequence 4, Appli
164	8	0.6	326	4	US-09-721-137-4	Sequence 4, Appli
165	8	0.6	342	4	US-09-118-464-3	Sequence 3, Appli
166	8	0.6	361	1	US-08-415-751-3	Sequence 3, Appli
167	8	0.6	361	4	US-09-404-296B-30	Sequence 30, Appl
168	8	0.6	372	4	US-09-118-464-2	Sequence 2, Appli
169	8	0.6	406	1	US-07-973-431B-1	Sequence 1, Appli
170	8	0.6	413	2	US-08-481-814A-8	Sequence 8, Appli
171	8	0.6	413	3	US-08-836-582-2	Sequence 2, Appli
172	8	0.6	413	4	US-09-265-566-2	Sequence 2, Appli
173	8	0.6	413	4	US-09-242-737-4	Sequence 4, Appli
174	8	0.6	444	1	US-07-881-075-3	Sequence 3, Appli
175	8	0.6	444	1	US-08-120-827-3	Sequence 3, Appli
176	8	0.6	444	1	US-08-478-675-3	Sequence 3, Appli
177	8	0.6	446	4	US-09-004-393B-4	Sequence 4, Appli
178	8	0.6	461	1	US-08-186-222-2	Sequence 2, Appli
179	8	0.6	478	3	US-08-155-888-2	Sequence 2, Appli
180	8	0.6	554	3	US-09-319-989-6	Sequence 6, Appli
181	8	0.6	586	3	US-08-419-810-12	Sequence 12, Appl
182	8	0.6	586	6	5405775-11	Patent No. 5405775
183	8	0.6	652	4	US-08-559-896B-2	Sequence 2, Appli
184	8	0.6	653	4	US-09-513-057C-13	Sequence 13, Appl
185	8	0.6	673	4	US-09-107-532A-5134	Sequence 5134, Ap
186	8	0.6	674	4	US-08-653-648A-14	Sequence 14, Appl
187	8	0.6	678	5	PCT-US93-03027-3	Sequence 3, Appli
188	8	0.6	787	4	US-09-721-383-2	Sequence 2, Appli
189	8	0.6	787	4	US-09-721-137-2	Sequence 2, Appli
190	8	0.6	834	2	US-08-861-464-4	Sequence 4, Appli
191	8	0.6	834	2	US-08-396-001-4	Sequence 4, Appli
192	8	0.6	834	3	US-09-323-433A-4	Sequence 4, Appli
193	8	0.6	984	1	US-08-257-073-3	Sequence 3, Appli
194	8	0.6	984	2	US-08-184-009-120	Sequence 120, App
195	8	0.6	984	2	US-08-458-356-120	Sequence 120, App
196	8	0.6	984	3	US-08-460-736-120	Sequence 120, App
197	8	0.6	984	4	US-09-535-370-120	Sequence 120, App
198	8	0.6	1242	4	US-09-508-691-1	Sequence 1, Appli
199	8	0.6	1243	2	US-08-557-139-2	Sequence 2, Appli
200	8	0.6	1284	4	US-09-343-494-9	Sequence 9, Appli
201	8	0.6	1284	4	US-09-358-383C-11	Sequence 11, Appl
202	8	0.6	1321	2	US-08-317-310A-64	Sequence 64, Appl
203	8	0.6	1850	4	US-09-620-093A-5	Sequence 5, Appli
204	8	0.6	2314	4	US-09-268-347-49	Sequence 49, Appl
205	8	0.6	2391	2	US-08-446-855A-2	Sequence 2, Appli
206	8	0.6	2391	3	US-09-150-741-2	Sequence 2, Appli
207	7	0.5	9	1	US-08-106-981-11	Sequence 11, Appl
208	7	0.5	14	1	US-08-346-293-17	Sequence 17, Appl
209	7	0.5	14	2	US-08-727-688-23	Sequence 23, Appl
210	7	0.5	18	4	US-08-248-058-1	Sequence 1, Appli
211	7	0.5	24	2	US-08-752-891-7	Sequence 7, Appli
212	7	0.5	24	2	US-09-144-178-7	Sequence 7, Appli
213	7	0.5	24	3	US-09-406-854-7	Sequence 7, Appli
214	7	0.5	24	3	US-09-098-901-6	Sequence 6, Appli
215	7	0.5	43	4	US-07-757-022B-12	Sequence 12, Appl
216	7	0.5	51	4	US-09-314-268-100	Sequence 100, App
217	7	0.5	69	2	US-08-583-569-1	Sequence 1, Appli
218	7	0.5	73	4	US-09-328-352-4184	Sequence 4184, Ap
219	7	0.5	75	3	US-08-927-219-51	Sequence 51, Appl
220	7	0.5	85	4	US-09-134-001C-3387	Sequence 3387, Ap
221	7	0.5	101	4	US-09-858-664A-10	Sequence 10, Appl
222	7	0.5	110	2	US-08-569-166-34	Sequence 34, Appl
223	7	0.5	110	3	US-08-961-083-102	Sequence 102, App
224	7	0.5	110	4	US-09-536-784-102	Sequence 102, App
225	7	0.5	111	4	US-09-858-664A-8	Sequence 8, Appli
226	7	0.5	129	4	US-09-328-352-6408	Sequence 6408, Ap
227	7	0.5	130	2	US-08-630-822A-90	Sequence 90, Appl
228	7	0.5	130	2	US-09-005-069-90	Sequence 90, Appl
229	7	0.5	130	4	US-09-171-156A-39	Sequence 39, Appl
230	7	0.5	130	4	US-09-004-730A-39	Sequence 39, Appl
231	7	0.5	130	4	US-08-981-799A-39	Sequence 39, Appl
232	7	0.5	148	4	US-09-134-001C-4679	Sequence 4679, Ap
233	7	0.5	148	4	US-09-461-325-453	Sequence 453, App
234	7	0.5	159	3	US-08-992-176-6	Sequence 6, Appli
235	7	0.5	159	4	US-09-107-532A-6434	Sequence 6434, Ap
236	7	0.5	172	4	US-07-757-022B-88	Sequence 88, Appl
237	7	0.5	179	4	US-09-252-991A-21165	Sequence 21165, A
238	7	0.5	179	4	US-09-198-452A-1064	Sequence 1064, Ap
239	7	0.5	180	4	US-09-134-001C-2966	Sequence 2966, Ap
240	7	0.5	187	4	US-09-247-155-102	Sequence 102, App
241	7	0.5	188	3	US-09-130-663-30	Sequence 30, Appl
242	7	0.5	188	3	US-09-332-934-14	Sequence 14, Appl
243	7	0.5	188	3	US-09-432-335-30	Sequence 30, Appl
244	7	0.5	188	4	US-09-614-022-30	Sequence 30, Appl
245	7	0.5	192	4	US-07-757-022B-90	Sequence 90, Appl
246	7	0.5	204	4	US-07-757-022B-92	Sequence 92, Appl

247	7	0.5	206	3	US-09-156-580-2	Sequence 2, Appli
248	7	0.5	207	4	US-09-199-637A-211	Sequence 211, App
249	7	0.5	208	4	US-07-757-022B-132	Sequence 132, App
250	7	0.5	209	4	US-07-757-022B-94	Sequence 94, Appl
251	7	0.5	211	2	US-08-708-958-2	Sequence 2, Appli
252	7	0.5	211	4	US-09-302-769-10	Sequence 10, Appl
253	7	0.5	211	4	US-09-972-800A-12	Sequence 12, Appl
254	7	0.5	212	4	US-09-302-769-4	Sequence 4, Appli
255	7	0.5	212	4	US-09-302-769-12	Sequence 12, Appl
256	7	0.5	212	4	US-09-399-913-63	Sequence 63, Appl
257	7	0.5	212	4	US-09-341-349D-2	Sequence 2, Appli
258	7	0.5	217	4	US-07-757-022B-76	Sequence 76, Appl
259	7	0.5	218	1	US-08-463-115-92	Sequence 92, Appl
260	7	0.5	218	1	US-08-465-388-92	Sequence 92, Appl
261	7	0.5	218	3	US-09-068-655-7	Sequence 7, Appli
262	7	0.5	219	1	US-08-463-115-91	Sequence 91, Appl
263	7	0.5	219	1	US-08-465-388-91	Sequence 91, Appl
264	7	0.5	219	4	US-09-134-001C-4581	Sequence 4581, Ap
265	7	0.5	220	4	US-07-757-022B-96	Sequence 96, Appl
266	7	0.5	223	4	US-09-206-676C-1	Sequence 1, Appli
267	7	0.5	223	4	US-09-009-816-4	Sequence 4, Appli
268	7	0.5	226	4	US-09-206-676C-2	Sequence 2, Appli
269	7	0.5	229	4	US-09-604-978-9	Sequence 9, Appli
270	7	0.5	229	4	US-09-604-728-9	Sequence 9, Appli
271	7	0.5	230	3	US-09-248-335-44	Sequence 44, Appl
272	7	0.5	231	4	US-07-757-022B-30	Sequence 30, Appl
273	7	0.5	233	2	US-08-458-568A-4	Sequence 4, Appli
274	7	0.5	243	4	US-09-166-350-19	Sequence 19, Appl
275	7	0.5	246	3	US-09-185-160-7	Sequence 7, Appli
276	7	0.5	250	2	US-08-861-269-5	Sequence 5, Appli
277	7	0.5	250	2	US-09-134-596-5	Sequence 5, Appli
278	7	0.5	250	3	US-09-293-273-5	Sequence 5, Appli
279	7	0.5	254	4	US-09-294-531B-3	Sequence 3, Appli
280	7	0.5	258	2	US-08-720-258-4	Sequence 4, Appli
281	7	0.5	262	1	US-08-106-981-6	Sequence 6, Appli
282	7	0.5	264	1	US-08-463-115-93	Sequence 93, Appl
283	7	0.5	264	1	US-08-465-388-93	Sequence 93, Appl
284	7	0.5	264	3	US-08-894-731-4	Sequence 4, Appli
285	7	0.5	266	4	US-09-252-991A-17646	Sequence 17646, A
286	7	0.5	285	3	US-08-482-085B-20	Sequence 20, Appl
287	7	0.5	288	4	US-09-134-001C-3292	Sequence 3292, Ap
288	7	0.5	296	4	US-07-757-022B-70	Sequence 70, Appl
289	7	0.5	302	4	US-09-725-311-2	Sequence 2, Appli
290	7	0.5	310	3	US-08-651-136C-22	Sequence 22, Appl
291	7	0.5	310	4	US-09-229-911A-22	Sequence 22, Appl
292	7	0.5	314	4	US-09-544-618-20	Sequence 20, Appl
293	7	0.5	315	4	US-09-404-296B-8	Sequence 8, Appli
294	7	0.5	320	4	US-09-134-001C-4439	Sequence 4439, Ap
295	7	0.5	325	4	US-09-252-991A-27128	Sequence 27128, A
296	7	0.5	330	4	US-09-725-311-4	Sequence 4, Appli
297	7	0.5	338	4	US-09-634-957-4	Sequence 4, Appli
298	7	0.5	338	4	US-09-723-153-4	Sequence 4, Appli
299	7	0.5	338	4	US-09-723-429-4	Sequence 4, Appli
300	7	0.5	345	2	US-08-282-197C-50	Sequence 50, Appl
301	7	0.5	346	4	US-09-724-224-6	Sequence 6, Appli
302	7	0.5	347	3	US-09-094-557-3	Sequence 3, Appli
303	7	0.5	349	4	US-09-162-524-3	Sequence 3, Appli
304	7	0.5	351	4	US-09-399-913-61	Sequence 61, Appl
305	7	0.5	351	4	US-09-252-991A-17990	Sequence 17990, A
306	7	0.5	360	3	US-08-899-437-7	Sequence 7, Appli
307	7	0.5	360	3	US-09-126-121-7	Sequence 7, Appli
308	7	0.5	366	4	US-09-718-692-4	Sequence 4, Appli
309	7	0.5	366	4	US-09-718-852-4	Sequence 4, Appli
310	7	0.5	366	4	US-09-718-815-4	Sequence 4, Appli
311	7	0.5	367	1	US-07-737-736B-4	Sequence 4, Appli
312	7	0.5	367	4	US-08-213-419B-15	Sequence 15, Appl
313	7	0.5	367	4	US-09-009-816-2	Sequence 2, Appli
314	7	0.5	368	3	US-09-413-574-4	Sequence 4, Appli
315	7	0.5	368	4	US-09-724-519-10	Sequence 10, Appl
316	7	0.5	368	4	US-09-592-037-10	Sequence 10, Appl
317	7	0.5	369	2	US-08-773-870-4	Sequence 4, Appli
318	7	0.5	370	4	US-09-724-224-2	Sequence 2, Appli
319	7	0.5	372	4	US-07-757-022B-64	Sequence 64, Appl
320	7	0.5	377	1	US-08-525-697-2	Sequence 2, Appli
321	7	0.5	382	4	US-08-213-419B-19	Sequence 19, Appl
322	7	0.5	382	4	US-09-724-519-4	Sequence 4, Appli
323	7	0.5	382	4	US-09-592-037-4	Sequence 4, Appli
324	7	0.5	391	1	US-08-602-010A-6	Sequence 6, Appli
325	7	0.5	391	1	US-08-680-726A-6	Sequence 6, Appli
326	7	0.5	391	3	US-09-092-409-6	Sequence 6, Appli
327	7	0.5	393	4	US-09-230-371A-29	Sequence 29, Appl
328	7	0.5	413	2	US-08-282-197C-49	Sequence 49, Appl
329	7	0.5	414	1	US-07-667-276A-4	Sequence 4, Appli
330	7	0.5	422	3	US-08-872-979-8	Sequence 8, Appli
331	7	0.5	422	4	US-07-757-022B-68	Sequence 68, Appl
332	7	0.5	423	4	US-07-757-022B-66	Sequence 66, Appl
333	7	0.5	433	4	US-09-595-424-2	Sequence 2, Appli
334	7	0.5	446	4	US-09-286-981B-6	Sequence 6, Appli
335	7	0.5	446	4	US-09-286-981B-9	Sequence 9, Appli
336	7	0.5	450	2	US-08-861-464-2	Sequence 2, Appli
337	7	0.5	450	2	US-08-396-001-2	Sequence 2, Appli
338	7	0.5	450	3	US-09-323-433A-2	Sequence 2, Appli
339	7	0.5	460	4	US-09-134-001C-5322	Sequence 5322, Ap
340	7	0.5	463	4	US-07-757-022B-54	Sequence 54, Appl
341	7	0.5	463	4	US-09-345-236B-2	Sequence 2, Appli
342	7	0.5	465	3	US-09-537-357-52	Sequence 52, Appl
343	7	0.5	465	4	US-09-252-991A-17635	Sequence 17635, A
344	7	0.5	467	4	US-08-495-484-12	Sequence 12, Appl
345	7	0.5	471	1	US-08-176-427B-11	Sequence 11, Appl
346	7	0.5	471	2	US-08-356-060A-34	Sequence 34, Appl
347	7	0.5	471	3	US-08-460-900C-34	Sequence 34, Appl
348	7	0.5	471	3	US-08-757-230A-8	Sequence 8, Appli
349	7	0.5	471	3	US-08-674-509B-34	Sequence 34, Appl
350	7	0.5	471	3	US-08-954-698-34	Sequence 34, Appl
351	7	0.5	471	4	US-08-957-874-34	Sequence 34, Appl
352	7	0.5	471	4	US-08-700-393-8	Sequence 8, Appli
353	7	0.5	471	4	US-09-639-695-34	Sequence 34, Appl
354	7	0.5	474	4	US-09-702-705-1812	Sequence 1812, Ap
355	7	0.5	474	4	US-09-736-457-1812	Sequence 1812, Ap
356	7	0.5	475	2	US-08-861-464-14	Sequence 14, Appl
357	7	0.5	475	2	US-08-396-001-14	Sequence 14, Appl
358	7	0.5	475	3	US-09-323-433A-14	Sequence 14, Appl
359	7	0.5	475	4	US-09-569-804-4	Sequence 4, Appli
360	7	0.5	476	3	US-09-316-083-3	Sequence 3, Appli
361	7	0.5	476	4	US-09-933-700-3	Sequence 3, Appli
362	7	0.5	484	4	US-09-242-913B-17	Sequence 17, Appl
363	7	0.5	487	4	US-09-724-224-8	Sequence 8, Appli
364	7	0.5	489	1	US-07-903-103-4	Sequence 4, Appli
365	7	0.5	489	1	US-08-044-619A-4	Sequence 4, Appli
366	7	0.5	489	1	US-08-283-911-4	Sequence 4, Appli
367	7	0.5	489	1	US-08-245-500A-5	Sequence 5, Appli
368	7	0.5	489	1	US-08-390-546-5	Sequence 5, Appli
369	7	0.5	489	1	US-08-390-479A-5	Sequence 5, Appli
370	7	0.5	489	1	US-08-557-393-5	Sequence 5, Appli
371	7	0.5	489	1	US-08-390-516C-5	Sequence 5, Appli
372	7	0.5	489	1	US-08-390-517A-5	Sequence 5, Appli
373	7	0.5	489	1	US-08-390-515A-5	Sequence 5, Appli
374	7	0.5	489	2	US-08-801-718-5	Sequence 5, Appli
375	7	0.5	489	4	US-09-170-159A-5	Sequence 5, Appli
376	7	0.5	489	4	US-09-480-718-46	Sequence 46, Appl
377	7	0.5	490	4	US-09-336-643A-6	Sequence 6, Appli
378	7	0.5	494	4	US-09-595-424-4	Sequence 4, Appli
379	7	0.5	496	3	US-09-350-268-2	Sequence 2, Appli
380	7	0.5	496	4	US-09-558-679-2	Sequence 2, Appli
381	7	0.5	500	4	US-09-325-932A-149	Sequence 149, App
382	7	0.5	506	2	US-08-820-170A-19	Sequence 19, Appl
383	7	0.5	506	3	US-09-055-699-19	Sequence 19, Appl
384	7	0.5	506	3	US-09-273-565-19	Sequence 19, Appl
385	7	0.5	506	4	US-09-565-538-19	Sequence 19, Appl
386	7	0.5	506	4	US-09-661-468-19	Sequence 19, Appl
387	7	0.5	506	4	US-09-976-165-19	Sequence 19, Appl
388	7	0.5	512	4	US-09-724-224-4	Sequence 4, Appli
389	7	0.5	513	4	US-09-724-519-6	Sequence 6, Appli
390	7	0.5	513	4	US-09-592-037-6	Sequence 6, Appli
391	7	0.5	516	1	US-08-097-829-4	Sequence 4, Appli
392	7	0.5	516	1	US-08-577-403-4	Sequence 4, Appli

393	7	0.5	524	4	US-09-186-276B-56	Sequence 56, Appl
394	7	0.5	524	4	US-08-842-445-56	Sequence 56, Appl
395	7	0.5	524	4	US-09-186-188B-56	Sequence 56, Appl
396	7	0.5	527	4	US-09-134-001C-3358	Sequence 3358, Ap
397	7	0.5	535	4	US-09-134-001C-3338	Sequence 3338, Ap
398	7	0.5	536	4	US-09-463-712C-10	Sequence 10, Appl
399	7	0.5	543	2	US-08-224-482-4	Sequence 4, Appli
400	7	0.5	543	2	US-08-469-412A-7	Sequence 7, Appli
401	7	0.5	543	3	US-09-021-715-7	Sequence 7, Appli
402	7	0.5	543	4	US-09-702-705-337	Sequence 337, App
403	7	0.5	543	4	US-09-736-457-337	Sequence 337, App
404	7	0.5	548	2	US-08-469-412A-2	Sequence 2, Appli
405	7	0.5	548	3	US-09-021-715-2	Sequence 2, Appli
406	7	0.5	551	4	US-09-252-991A-27787	Sequence 27787, A
407	7	0.5	553	4	US-09-242-913B-13	Sequence 13, Appl
408	7	0.5	568	1	US-08-320-559-30	Sequence 30, Appl
409	7	0.5	568	3	US-08-545-860D-30	Sequence 30, Appl
410	7	0.5	568	5	PCT-US94-04496-30	Sequence 30, Appl
411	7	0.5	575	4	US-09-724-519-8	Sequence 8, Appli
412	7	0.5	575	4	US-09-592-037-8	Sequence 8, Appli
413	7	0.5	579	4	US-09-529-279-4	Sequence 4, Appli
414	7	0.5	579	4	US-10-158-895-4	Sequence 4, Appli
415	7	0.5	582	3	US-08-906-865-3	Sequence 3, Appli
416	7	0.5	582	4	US-09-428-711A-2	Sequence 2, Appli
417	7	0.5	582	4	US-09-129-668-3	Sequence 3, Appli
418	7	0.5	590	2	US-08-785-310A-5	Sequence 5, Appli
419	7	0.5	590	4	US-09-529-279-15	Sequence 15, Appl
420	7	0.5	590	4	US-10-158-895-15	Sequence 15, Appl
421	7	0.5	592	1	US-08-217-327-8	Sequence 8, Appli
422	7	0.5	594	2	US-08-785-310A-6	Sequence 6, Appli
423	7	0.5	607	4	US-09-252-991A-32643	Sequence 32643, A
424	7	0.5	609	4	US-09-198-452A-579	Sequence 579, App
425	7	0.5	637	4	US-09-284-768A-20	Sequence 20, Appl
426	7	0.5	646	4	US-09-328-352-6017	Sequence 6017, Ap
427	7	0.5	658	4	US-09-595-424-6	Sequence 6, Appli
428	7	0.5	660	3	US-09-058-489-18	Sequence 18, Appl
429	7	0.5	660	3	US-09-058-489-91	Sequence 91, Appl
430	7	0.5	662	3	US-09-058-489-15	Sequence 15, Appl
431	7	0.5	662	3	US-09-058-489-16	Sequence 16, Appl
432	7	0.5	665	4	US-09-595-684B-35	Sequence 35, Appl
433	7	0.5	674	4	US-09-284-768A-21	Sequence 21, Appl
434	7	0.5	693	4	US-09-107-532A-5812	Sequence 5812, Ap
435	7	0.5	694	3	US-08-559-397A-31	Sequence 31, Appl
436	7	0.5	696	3	US-08-899-437-23	Sequence 23, Appl
437	7	0.5	696	3	US-09-126-121-23	Sequence 23, Appl
438	7	0.5	703	3	US-08-910-925-4	Sequence 4, Appli
439	7	0.5	706	3	US-09-408-820-2	Sequence 2, Appli
440	7	0.5	706	4	US-09-914-259-29	Sequence 29, Appl
441	7	0.5	715	3	US-08-669-286-5	Sequence 5, Appli
442	7	0.5	715	3	US-09-469-253-5	Sequence 5, Appli
443	7	0.5	715	3	US-09-642-146-5	Sequence 5, Appli
444	7	0.5	716	4	US-09-107-532A-5208	Sequence 5208, Ap
445	7	0.5	717	3	US-08-910-925-1	Sequence 1, Appli
446	7	0.5	720	3	US-08-899-437-6	Sequence 6, Appli
447	7	0.5	720	3	US-09-126-121-6	Sequence 6, Appli
448	7	0.5	731	3	US-09-185-160-11	Sequence 11, Appl
449	7	0.5	734	3	US-09-185-160-13	Sequence 13, Appl
450	7	0.5	739	4	US-09-134-001C-3586	Sequence 3586, Ap
451	7	0.5	743	3	US-08-910-925-3	Sequence 3, Appli
452	7	0.5	750	3	US-09-185-160-14	Sequence 14, Appl
453	7	0.5	752	1	US-08-244-189-2	Sequence 2, Appli
454	7	0.5	764	4	US-09-370-838-67	Sequence 67, Appl
455	7	0.5	789	3	US-08-727-308-1	Sequence 1, Appli
456	7	0.5	789	4	US-09-390-234-16	Sequence 16, Appl
457	7	0.5	790	4	US-09-634-957-2	Sequence 2, Appli
458	7	0.5	790	4	US-09-723-153-2	Sequence 2, Appli
459	7	0.5	790	4	US-09-723-429-2	Sequence 2, Appli
460	7	0.5	805	4	US-09-425-335-6	Sequence 6, Appli
461	7	0.5	809	4	US-09-186-276B-58	Sequence 58, Appl
462	7	0.5	809	4	US-08-842-445-58	Sequence 58, Appl
463	7	0.5	809	4	US-09-186-188B-58	Sequence 58, Appl
464	7	0.5	821	1	US-09-377-465A-2	Sequence 2, Appli
465	7	0.5	823	4	US-09-651-656-19	Sequence 19, Appl
466	7	0.5	823	4	US-09-650-855-19	Sequence 19, Appl
467	7	0.5	845	4	US-09-198-452A-458	Sequence 458, App
468	7	0.5	846	2	US-07-728-215-33	Sequence 33, Appl
469	7	0.5	846	4	US-08-938-085A-33	Sequence 33, Appl
470	7	0.5	846	4	US-10-072-844-33	Sequence 33, Appl
471	7	0.5	866	2	US-08-483-101-4	Sequence 4, Appli
472	7	0.5	878	4	US-09-735-934A-2	Sequence 2, Appli
473	7	0.5	878	4	US-10-060-332-2	Sequence 2, Appli
474	7	0.5	898	4	US-09-585-858-37	Sequence 37, Appl
475	7	0.5	928	4	US-09-914-259-23	Sequence 23, Appl
476	7	0.5	947	4	US-09-418-780A-1	Sequence 1, Appli
477	7	0.5	957	4	US-09-595-684B-27	Sequence 27, Appl
478	7	0.5	957	4	US-09-107-532A-4021	Sequence 4021, Ap
479	7	0.5	960	4	US-09-595-424-8	Sequence 8, Appli
480	7	0.5	989	4	US-08-213-419B-2	Sequence 2, Appli
481	7	0.5	989	4	US-08-213-419B-4	Sequence 4, Appli
482	7	0.5	1003	4	US-09-521-511C-11	Sequence 11, Appl
483	7	0.5	1049	4	US-07-757-022B-58	Sequence 58, Appl
484	7	0.5	1050	4	US-09-428-711A-16	Sequence 16, Appl
485	7	0.5	1051	4	US-09-428-711A-14	Sequence 14, Appl
486	7	0.5	1053	4	US-09-724-519-2	Sequence 2, Appli
487	7	0.5	1053	4	US-09-592-037-2	Sequence 2, Appli
488	7	0.5	1056	4	US-09-595-684B-29	Sequence 29, Appl
489	7	0.5	1057	3	US-09-541-782-10	Sequence 10, Appl
490	7	0.5	1057	4	US-09-723-820-10	Sequence 10, Appl
491	7	0.5	1066	3	US-09-541-782-8	Sequence 8, Appli
492	7	0.5	1066	4	US-09-723-820-8	Sequence 8, Appli
493	7	0.5	1073	3	US-09-541-782-6	Sequence 6, Appli
494	7	0.5	1073	4	US-09-723-820-6	Sequence 6, Appli
495	7	0.5	1093	3	US-08-545-860D-55	Sequence 55, Appl
496	7	0.5	1093	5	PCT-US94-04496-55	Sequence 55, Appl
497	7	0.5	1103	3	US-09-162-373-1	Sequence 1, Appli
498	7	0.5	1103	3	US-09-467-946-1	Sequence 1, Appli
499	7	0.5	1111	4	US-09-914-259-28	Sequence 28, Appl
500	7	0.5	1140	4	US-07-757-022B-104	Sequence 104, App
501	7	0.5	1141	1	US-08-131-365B-54	Sequence 54, Appl
502	7	0.5	1141	2	US-08-668-123-54	Sequence 54, Appl
503	7	0.5	1142	2	US-08-993-118-7	Sequence 7, Appli
504	7	0.5	1142	3	US-08-845-528C-7	Sequence 7, Appli
505	7	0.5	1142	3	US-09-061-709-2	Sequence 2, Appli
506	7	0.5	1142	4	US-09-066-281B-7	Sequence 7, Appli
507	7	0.5	1142	4	US-09-899-651-2	Sequence 2, Appli
508	7	0.5	1184	3	US-09-541-782-2	Sequence 2, Appli
509	7	0.5	1184	4	US-09-723-820-2	Sequence 2, Appli
510	7	0.5	1184	4	US-09-266-225D-18	Sequence 18, Appl
511	7	0.5	1185	3	US-09-041-886-23	Sequence 23, Appl
512	7	0.5	1187	1	US-08-320-559-28	Sequence 28, Appl
513	7	0.5	1187	3	US-08-545-860D-28	Sequence 28, Appl
514	7	0.5	1187	5	PCT-US94-04496-28	Sequence 28, Appl
515	7	0.5	1210	1	US-08-320-559-26	Sequence 26, Appl
516	7	0.5	1210	3	US-08-545-860D-26	Sequence 26, Appl
517	7	0.5	1210	5	PCT-US94-04496-26	Sequence 26, Appl
518	7	0.5	1235	1	US-08-118-101A-2	Sequence 2, Appli
519	7	0.5	1298	2	US-08-690-473-2	Sequence 2, Appli
520	7	0.5	1298	3	US-09-259-821A-2	Sequence 2, Appli
521	7	0.5	1298	3	US-08-843-659-2	Sequence 2, Appli
522	7	0.5	1313	4	US-07-757-022B-142	Sequence 142, App
523	7	0.5	1314	4	US-07-757-022B-50	Sequence 50, Appl
524	7	0.5	1328	3	US-08-781-891-76	Sequence 76, Appl
525	7	0.5	1328	4	US-09-618-166-76	Sequence 76, Appl
526	7	0.5	1354	4	US-07-757-022B-48	Sequence 48, Appl
527	7	0.5	1363	4	US-07-757-022B-52	Sequence 52, Appl
528	7	0.5	1404	4	US-07-757-022B-2	Sequence 2, Appli
529	7	0.5	1404	4	US-07-757-022B-62	Sequence 62, Appl
530	7	0.5	1451	4	US-09-060-299-25	Sequence 25, Appl
531	7	0.5	1451	4	US-09-402-923A-25	Sequence 25, Appl
532	7	0.5	1481	2	US-08-616-844-40	Sequence 40, Appl
533	7	0.5	1481	2	US-08-599-654-40	Sequence 40, Appl
534	7	0.5	1481	3	US-08-944-868A-40	Sequence 40, Appl
535	7	0.5	1481	3	US-08-944-423A-40	Sequence 40, Appl
536	7	0.5	1481	3	US-08-944-496-40	Sequence 40, Appl
537	7	0.5	1581	3	US-09-110-517-2	Sequence 2, Appli
538	7	0.5	1584	4	US-09-060-299-39	Sequence 39, Appl

539	7	0.5	1584	4	US-09-402-923A-39	Sequence 39, Appl
540	7	0.5	1586	4	US-09-060-299-44	Sequence 44, Appl
541	7	0.5	1586	4	US-09-402-923A-44	Sequence 44, Appl
542	7	0.5	1591	4	US-09-060-299-4	Sequence 4, Appli
543	7	0.5	1591	4	US-09-060-299-43	Sequence 43, Appl
544	7	0.5	1591	4	US-09-402-923A-4	Sequence 4, Appli
545	7	0.5	1591	4	US-09-402-923A-43	Sequence 43, Appl
546	7	0.5	1614	4	US-09-060-299-42	Sequence 42, Appl
547	7	0.5	1614	4	US-09-402-923A-42	Sequence 42, Appl
548	7	0.5	1615	4	US-09-060-299-3	Sequence 3, Appli
549	7	0.5	1615	4	US-09-402-923A-3	Sequence 3, Appli
550	7	0.5	1622	4	US-09-231-899-72	Sequence 72, Appl
551	7	0.5	1627	1	US-07-665-792E-9	Sequence 9, Appli
552	7	0.5	1637	4	US-09-718-692-2	Sequence 2, Appli
553	7	0.5	1637	4	US-09-718-852-2	Sequence 2, Appli
554	7	0.5	1637	4	US-09-718-815-2	Sequence 2, Appli
555	7	0.5	1639	4	US-09-060-299-29	Sequence 29, Appl
556	7	0.5	1639	4	US-09-402-923A-29	Sequence 29, Appl
557	7	0.5	1665	4	US-09-858-664A-2	Sequence 2, Appli
558	7	0.5	1690	4	US-09-595-684B-39	Sequence 39, Appl
559	7	0.5	1706	2	US-08-459-568-2	Sequence 2, Appli
560	7	0.5	1706	2	US-08-399-411-2	Sequence 2, Appli
561	7	0.5	1706	3	US-08-516-859A-2	Sequence 2, Appli
562	7	0.5	1706	4	US-09-586-472-2	Sequence 2, Appli
563	7	0.5	1706	4	US-09-528-706-2	Sequence 2, Appli
564	7	0.5	1719	2	US-08-459-568-4	Sequence 4, Appli
565	7	0.5	1719	2	US-08-399-411-4	Sequence 4, Appli
566	7	0.5	1719	3	US-08-516-859A-4	Sequence 4, Appli
567	7	0.5	1719	4	US-09-586-472-4	Sequence 4, Appli
568	7	0.5	1719	4	US-09-528-706-4	Sequence 4, Appli
569	7	0.5	1783	4	US-09-362-336A-2	Sequence 2, Appli
570	7	0.5	1804	4	US-09-362-336A-4	Sequence 4, Appli
571	7	0.5	1805	4	US-09-004-838-92	Sequence 92, Appl
572	7	0.5	1817	4	US-09-004-838-125	Sequence 125, App
573	7	0.5	1854	4	US-09-004-838-108	Sequence 108, App
574	7	0.5	1874	4	US-09-331-403-2	Sequence 2, Appli
575	7	0.5	1890	4	US-09-004-838-88	Sequence 88, Appl
576	7	0.5	2265	2	US-08-149-097D-36	Sequence 36, Appl
577	7	0.5	2509	2	US-08-149-097D-35	Sequence 35, Appl
578	7	0.5	2662	4	US-09-595-684B-31	Sequence 31, Appl
579	7	0.5	3788	4	US-09-336-447A-76	Sequence 76, Appl
580	7	0.5	3969	3	US-08-061-376-5	Sequence 5, Appli
581	6	0.5	7	3	US-08-478-316-51	Sequence 51, Appl
582	6	0.5	7	4	US-09-019-793A-51	Sequence 51, Appl
583	6	0.5	9	4	US-09-144-280-13	Sequence 13, Appl
584	6	0.5	10	4	US-09-468-265-6	Sequence 6, Appli
585	6	0.5	10	4	US-09-434-476A-6	Sequence 6, Appli
586	6	0.5	10	4	US-09-434-476A-7	Sequence 7, Appli
587	6	0.5	12	4	US-08-481-968A-27	Sequence 27, Appl
588	6	0.5	12	4	US-08-154-712B-27	Sequence 27, Appl
589	6	0.5	12	4	US-09-257-825B-26	Sequence 26, Appl
590	6	0.5	13	1	US-08-346-293-16	Sequence 16, Appl
591	6	0.5	13	1	US-08-057-167-16	Sequence 16, Appl
592	6	0.5	13	5	PCT-US93-05412-16	Sequence 16, Appl
593	6	0.5	14	6	5204097-4	Patent No. 5204097
594	6	0.5	15	4	US-08-481-968A-17	Sequence 17, Appl
595	6	0.5	15	4	US-08-154-712B-17	Sequence 17, Appl
596	6	0.5	16	3	US-08-851-843A-201	Sequence 201, App
597	6	0.5	16	3	US-08-974-549A-320	Sequence 320, App
598	6	0.5	16	3	US-08-931-858E-92	Sequence 92, Appl
599	6	0.5	16	3	US-08-981-739-92	Sequence 92, Appl
600	6	0.5	16	3	US-09-146-755-1	Sequence 1, Appli
601	6	0.5	16	3	US-08-854-050-201	Sequence 201, App
602	6	0.5	16	4	US-09-430-323-201	Sequence 201, App
603	6	0.5	16	4	US-09-128-026-92	Sequence 92, Appl
604	6	0.5	18	3	US-09-100-414B-26	Sequence 26, Appl
605	6	0.5	18	3	US-09-100-414B-30	Sequence 30, Appl
606	6	0.5	18	3	US-09-303-323-26	Sequence 26, Appl
607	6	0.5	18	3	US-09-303-323-30	Sequence 30, Appl
608	6	0.5	18	4	US-09-770-014-26	Sequence 26, Appl
609	6	0.5	18	4	US-09-770-014-30	Sequence 30, Appl
610	6	0.5	19	3	US-08-817-926-12	Sequence 12, Appl
611	6	0.5	20	6	5204097-3	Patent No. 5204097
612	6	0.5	22	3	US-08-940-095-173	Sequence 173, App
613	6	0.5	22	3	US-08-940-095-175	Sequence 175, App
614	6	0.5	22	3	US-08-940-093-173	Sequence 173, App
615	6	0.5	22	3	US-08-940-093-175	Sequence 175, App
616	6	0.5	22	3	US-08-940-096-173	Sequence 173, App
617	6	0.5	22	3	US-08-940-096-175	Sequence 175, App
618	6	0.5	22	3	US-09-465-719-173	Sequence 173, App
619	6	0.5	22	3	US-09-465-719-175	Sequence 175, App
620	6	0.5	22	4	US-09-453-605-173	Sequence 173, App
621	6	0.5	22	4	US-09-453-605-175	Sequence 175, App
622	6	0.5	22	4	US-09-453-838-173	Sequence 173, App
623	6	0.5	22	4	US-09-453-838-175	Sequence 175, App
624	6	0.5	22	4	US-08-940-136-173	Sequence 173, App
625	6	0.5	22	4	US-08-940-136-175	Sequence 175, App
626	6	0.5	22	4	US-09-453-841-173	Sequence 173, App
627	6	0.5	22	4	US-09-453-841-175	Sequence 175, App
628	6	0.5	24	3	US-09-098-901-4	Sequence 4, Appli
629	6	0.5	24	3	US-09-098-901-5	Sequence 5, Appli
630	6	0.5	24	6	5198345-26	Patent No. 5198345
631	6	0.5	25	2	US-08-473-475A-6	Sequence 6, Appli
632	6	0.5	25	3	US-09-200-757-5	Sequence 5, Appli
633	6	0.5	25	6	5198345-12	Patent No. 5198345
634	6	0.5	26	1	US-08-254-493-7	Sequence 7, Appli
635	6	0.5	26	1	US-08-408-222B-7	Sequence 7, Appli
636	6	0.5	27	3	US-09-056-226-8	Sequence 8, Appli
637	6	0.5	28	6	5198345-19	Patent No. 5198345
638	6	0.5	30	1	US-08-446-692-39	Sequence 39, Appl
639	6	0.5	30	2	US-08-488-351A-39	Sequence 39, Appl
640	6	0.5	30	3	US-09-100-414B-66	Sequence 66, Appl
641	6	0.5	30	3	US-09-100-414B-70	Sequence 70, Appl
642	6	0.5	30	3	US-09-303-323-66	Sequence 66, Appl
643	6	0.5	30	3	US-09-303-323-70	Sequence 70, Appl
644	6	0.5	30	4	US-09-770-014-66	Sequence 66, Appl
645	6	0.5	30	4	US-09-770-014-70	Sequence 70, Appl
646	6	0.5	31	3	US-09-023-082A-25	Sequence 25, Appl
647	6	0.5	34	4	US-09-301-978C-19	Sequence 19, Appl
648	6	0.5	35	3	US-09-082-279B-289	Sequence 289, App
649	6	0.5	35	3	US-09-082-279B-290	Sequence 290, App
650	6	0.5	35	3	US-09-082-279B-291	Sequence 291, App
651	6	0.5	35	3	US-09-082-279B-292	Sequence 292, App
652	6	0.5	35	3	US-09-082-279B-293	Sequence 293, App
653	6	0.5	35	3	US-09-082-279B-294	Sequence 294, App
654	6	0.5	35	3	US-09-082-279B-295	Sequence 295, App
655	6	0.5	35	3	US-09-082-279B-296	Sequence 296, App
656	6	0.5	35	3	US-09-082-279B-297	Sequence 297, App
657	6	0.5	35	3	US-09-082-279B-298	Sequence 298, App
658	6	0.5	35	3	US-09-082-279B-299	Sequence 299, App
659	6	0.5	35	3	US-09-082-279B-300	Sequence 300, App
660	6	0.5	35	3	US-09-082-279B-301	Sequence 301, App
661	6	0.5	35	3	US-09-082-279B-302	Sequence 302, App
662	6	0.5	35	3	US-09-082-279B-303	Sequence 303, App
663	6	0.5	35	3	US-09-082-279B-304	Sequence 304, App
664	6	0.5	35	3	US-09-082-279B-305	Sequence 305, App
665	6	0.5	35	3	US-09-082-279B-306	Sequence 306, App
666	6	0.5	35	4	US-09-315-304B-289	Sequence 289, App
667	6	0.5	35	4	US-09-315-304B-290	Sequence 290, App
668	6	0.5	35	4	US-09-315-304B-291	Sequence 291, App
669	6	0.5	35	4	US-09-315-304B-292	Sequence 292, App
670	6	0.5	35	4	US-09-315-304B-293	Sequence 293, App
671	6	0.5	35	4	US-09-315-304B-294	Sequence 294, App
672	6	0.5	35	4	US-09-315-304B-295	Sequence 295, App
673	6	0.5	35	4	US-09-315-304B-296	Sequence 296, App
674	6	0.5	35	4	US-09-315-304B-297	Sequence 297, App
675	6	0.5	35	4	US-09-315-304B-298	Sequence 298, App
676	6	0.5	35	4	US-09-315-304B-299	Sequence 299, App
677	6	0.5	35	4	US-09-315-304B-300	Sequence 300, App
678	6	0.5	35	4	US-09-315-304B-301	Sequence 301, App
679	6	0.5	35	4	US-09-315-304B-302	Sequence 302, App
680	6	0.5	35	4	US-09-315-304B-303	Sequence 303, App
681	6	0.5	35	4	US-09-315-304B-304	Sequence 304, App
682	6	0.5	35	4	US-09-315-304B-305	Sequence 305, App
683	6	0.5	35	4	US-09-315-304B-306	Sequence 306, App
684	6	0.5	35	4	US-09-834-784-289	Sequence 289, App

685	6	0.5	35	4	US-09-834-784-290	Sequence 290, App
686	6	0.5	35	4	US-09-834-784-291	Sequence 291, App
687	6	0.5	35	4	US-09-834-784-292	Sequence 292, App
688	6	0.5	35	4	US-09-834-784-293	Sequence 293, App
689	6	0.5	35	4	US-09-834-784-294	Sequence 294, App
690	6	0.5	35	4	US-09-834-784-295	Sequence 295, App
691	6	0.5	35	4	US-09-834-784-296	Sequence 296, App
692	6	0.5	35	4	US-09-834-784-297	Sequence 297, App
693	6	0.5	35	4	US-09-834-784-298	Sequence 298, App
694	6	0.5	35	4	US-09-834-784-299	Sequence 299, App
695	6	0.5	35	4	US-09-834-784-300	Sequence 300, App
696	6	0.5	35	4	US-09-834-784-301	Sequence 301, App
697	6	0.5	35	4	US-09-834-784-302	Sequence 302, App
698	6	0.5	35	4	US-09-834-784-303	Sequence 303, App
699	6	0.5	35	4	US-09-834-784-304	Sequence 304, App
700	6	0.5	35	4	US-09-834-784-305	Sequence 305, App
701	6	0.5	35	4	US-09-834-784-306	Sequence 306, App
702	6	0.5	37	2	US-08-631-328-45	Sequence 45, Appl
703	6	0.5	37	2	US-08-631-328-48	Sequence 48, Appl
704	6	0.5	37	2	US-08-631-328-52	Sequence 52, Appl
705	6	0.5	37	2	US-08-631-328-53	Sequence 53, Appl
706	6	0.5	37	4	US-09-732-210-510	Sequence 510, App
707	6	0.5	38	1	US-07-977-630-69	Sequence 69, Appl
708	6	0.5	45	3	US-08-776-059-11	Sequence 11, Appl
709	6	0.5	45	3	US-09-205-258-554	Sequence 554, App
710	6	0.5	46	4	US-09-369-247-127	Sequence 127, App
711	6	0.5	48	1	US-08-446-692-41	Sequence 41, Appl
712	6	0.5	48	2	US-08-488-351A-41	Sequence 41, Appl
713	6	0.5	49	2	US-08-248-839C-81	Sequence 81, Appl
714	6	0.5	51	1	US-08-188-228-32	Sequence 32, Appl
715	6	0.5	51	1	US-08-332-643-32	Sequence 32, Appl
716	6	0.5	51	1	US-08-332-638-32	Sequence 32, Appl
717	6	0.5	51	2	US-08-760-075A-27	Sequence 27, Appl
718	6	0.5	51	3	US-09-338-546-27	Sequence 27, Appl
719	6	0.5	51	4	US-09-659-084-27	Sequence 27, Appl
720	6	0.5	52	6	5198345-14	Patent No. 5198345
721	6	0.5	64	3	US-09-188-930-122	Sequence 122, App
722	6	0.5	64	4	US-09-312-283C-122	Sequence 122, App
723	6	0.5	64	4	US-09-107-532A-4650	Sequence 4650, Ap
724	6	0.5	65	4	US-09-230-041-28	Sequence 28, Appl
725	6	0.5	67	4	US-09-230-041-26	Sequence 26, Appl
726	6	0.5	69	2	US-08-292-968-21	Sequence 21, Appl
727	6	0.5	69	2	US-08-467-974-21	Sequence 21, Appl
728	6	0.5	69	2	US-08-467-536-21	Sequence 21, Appl
729	6	0.5	69	3	US-08-467-976-21	Sequence 21, Appl
730	6	0.5	69	3	US-08-995-159-3	Sequence 3, Appli
731	6	0.5	69	3	US-09-082-514-21	Sequence 21, Appl
732	6	0.5	69	4	US-09-134-001C-3787	Sequence 3787, Ap
733	6	0.5	69	4	US-09-545-605-3	Sequence 3, Appli
734	6	0.5	69	4	US-09-732-210-1003	Sequence 1003, Ap
735	6	0.5	69	4	US-09-732-210-1004	Sequence 1004, Ap
736	6	0.5	70	4	US-09-134-001C-4143	Sequence 4143, Ap
737	6	0.5	70	6	5204097-1	Patent No. 5204097
738	6	0.5	71	4	US-09-134-001C-4271	Sequence 4271, Ap
739	6	0.5	73	3	US-09-100-804-28	Sequence 28, Appl
740	6	0.5	75	3	US-08-480-640A-191	Sequence 191, App
741	6	0.5	75	3	US-08-686-968C-191	Sequence 191, App
742	6	0.5	75	3	US-08-488-237A-191	Sequence 191, App
743	6	0.5	75	4	US-08-375-992A-191	Sequence 191, App
744	6	0.5	75	4	US-08-472-679H-191	Sequence 191, App
745	6	0.5	76	4	US-09-328-352-4506	Sequence 4506, Ap
746	6	0.5	77	4	US-09-328-352-4362	Sequence 4362, Ap
747	6	0.5	79	2	US-08-530-290-16	Sequence 16, Appl
748	6	0.5	79	4	US-08-858-207A-305	Sequence 305, App
749	6	0.5	79	4	US-09-328-352-4679	Sequence 4679, Ap
750	6	0.5	81	4	US-09-134-001C-5248	Sequence 5248, Ap
751	6	0.5	81	4	US-09-198-452A-1186	Sequence 1186, Ap
752	6	0.5	86	4	US-09-134-001C-3986	Sequence 3986, Ap
753	6	0.5	88	4	US-09-134-001C-4055	Sequence 4055, Ap
754	6	0.5	89	3	US-08-816-977-4	Sequence 4, Appli
755	6	0.5	89	4	US-09-107-532A-6533	Sequence 6533, Ap
756	6	0.5	93	3	US-08-931-858E-51	Sequence 51, Appl
757	6	0.5	93	3	US-08-931-858E-134	Sequence 134, App
758	6	0.5	93	3	US-08-931-858E-176	Sequence 176, App
759	6	0.5	93	3	US-09-220-528-16	Sequence 16, Appl
760	6	0.5	93	4	US-08-469-260A-19	Sequence 19, Appl
761	6	0.5	93	4	US-08-488-446-19	Sequence 19, Appl
762	6	0.5	93	4	US-08-467-344A-19	Sequence 19, Appl
763	6	0.5	94	3	US-08-931-858E-87	Sequence 87, Appl
764	6	0.5	94	3	US-08-931-858E-172	Sequence 172, App
765	6	0.5	94	3	US-08-981-739-87	Sequence 87, Appl
766	6	0.5	94	3	US-08-981-739-172	Sequence 172, App
767	6	0.5	94	4	US-09-128-026-87	Sequence 87, Appl
768	6	0.5	94	4	US-09-128-026-172	Sequence 172, App
769	6	0.5	97	3	US-08-816-977-23	Sequence 23, Appl
770	6	0.5	98	4	US-09-134-001C-4699	Sequence 4699, Ap
771	6	0.5	103	3	US-08-535-681-46	Sequence 46, Appl
772	6	0.5	103	3	US-08-535-681-49	Sequence 49, Appl
773	6	0.5	103	4	US-09-615-192A-398	Sequence 398, App
774	6	0.5	104	3	US-08-535-681-44	Sequence 44, Appl
775	6	0.5	104	3	US-08-535-681-48	Sequence 48, Appl
776	6	0.5	105	4	US-08-311-731A-227	Sequence 227, App
777	6	0.5	106	6	5204258-4	Patent No. 5204258
778	6	0.5	108	4	US-09-107-532A-4307	Sequence 4307, Ap
779	6	0.5	110	4	US-09-732-210-299	Sequence 299, App
780	6	0.5	112	4	US-09-134-001C-3352	Sequence 3352, Ap
781	6	0.5	113	4	US-09-378-088A-92	Sequence 92, Appl
782	6	0.5	113	4	US-09-378-088A-94	Sequence 94, Appl
783	6	0.5	113	4	US-09-107-532A-4396	Sequence 4396, Ap
784	6	0.5	114	3	US-08-535-681-42	Sequence 42, Appl
785	6	0.5	114	3	US-08-535-681-50	Sequence 50, Appl
786	6	0.5	115	1	US-08-152-922A-7	Sequence 7, Appli
787	6	0.5	115	4	US-09-219-983A-4	Sequence 4, Appli
788	6	0.5	117	4	US-09-378-088A-96	Sequence 96, Appl
789	6	0.5	117	4	US-09-378-088A-98	Sequence 98, Appl
790	6	0.5	117	4	US-09-378-088A-100	Sequence 100, App
791	6	0.5	117	4	US-09-378-088A-102	Sequence 102, App
792	6	0.5	117	4	US-09-378-088A-104	Sequence 104, App
793	6	0.5	117	4	US-09-378-088A-106	Sequence 106, App
794	6	0.5	119	3	US-09-023-082A-80	Sequence 80, Appl
795	6	0.5	119	4	US-09-107-532A-3700	Sequence 3700, Ap
796	6	0.5	120	1	US-08-249-013-9	Sequence 9, Appli
797	6	0.5	120	2	US-08-886-863-9	Sequence 9, Appli
798	6	0.5	120	4	US-09-175-229-9	Sequence 9, Appli
799	6	0.5	120	5	PCT-US95-06764-9	Sequence 9, Appli
800	6	0.5	121	4	US-09-107-532A-7107	Sequence 7107, Ap
801	6	0.5	122	4	US-09-107-532A-5730	Sequence 5730, Ap
802	6	0.5	122	4	US-09-107-532A-6744	Sequence 6744, Ap
803	6	0.5	123	1	US-08-131-625B-17	Sequence 17, Appl
804	6	0.5	123	2	US-08-799-464A-13	Sequence 13, Appl
805	6	0.5	123	3	US-08-686-968C-11	Sequence 11, Appl
806	6	0.5	123	3	US-09-113-750A-42	Sequence 42, Appl
807	6	0.5	123	4	US-09-331-924-2	Sequence 2, Appli
808	6	0.5	123	4	US-09-732-210-209	Sequence 209, App
809	6	0.5	123	5	PCT-US95-09927-13	Sequence 13, Appl
810	6	0.5	123	5	PCT-US95-10904-19	Sequence 19, Appl
811	6	0.5	123	5	PCT-US95-10904-53	Sequence 53, Appl
812	6	0.5	123	5	PCT-US95-10904-55	Sequence 55, Appl
813	6	0.5	123	5	PCT-US95-10904-57	Sequence 57, Appl
814	6	0.5	123	5	PCT-US95-10904-59	Sequence 59, Appl
815	6	0.5	123	5	PCT-US95-10904-61	Sequence 61, Appl
816	6	0.5	126	4	US-09-328-352-7949	Sequence 7949, Ap
817	6	0.5	127	4	US-09-134-001C-5283	Sequence 5283, Ap
818	6	0.5	129	3	US-08-942-686-3	Sequence 3, Appli
819	6	0.5	129	3	US-08-942-686-4	Sequence 4, Appli
820	6	0.5	130	4	US-09-134-001C-3231	Sequence 3231, Ap
821	6	0.5	132	4	US-09-563-269-4	Sequence 4, Appli
822	6	0.5	132	4	US-09-732-210-123	Sequence 123, App
823	6	0.5	134	1	US-08-564-458-1	Sequence 1, Appli
824	6	0.5	134	1	US-08-564-833-1	Sequence 1, Appli
825	6	0.5	134	1	US-08-535-682-1	Sequence 1, Appli
826	6	0.5	134	1	US-08-446-383A-1	Sequence 1, Appli
827	6	0.5	134	1	US-08-446-383A-2	Sequence 2, Appli
828	6	0.5	134	1	US-08-795-628-1	Sequence 1, Appli
829	6	0.5	134	1	US-08-519-777-76	Sequence 76, Appl
830	6	0.5	134	1	US-08-519-777-77	Sequence 77, Appl

831	6	0.5	134	1	US-08-519-777-78	Sequence 78, Appl
832	6	0.5	134	1	US-08-618-543-1	Sequence 1, Appli
833	6	0.5	134	1	US-08-742-035-76	Sequence 76, Appl
834	6	0.5	134	1	US-08-742-035-77	Sequence 77, Appl
835	6	0.5	134	1	US-08-742-035-78	Sequence 78, Appl
836	6	0.5	134	2	US-08-777-019-76	Sequence 76, Appl
837	6	0.5	134	2	US-08-777-019-77	Sequence 77, Appl
838	6	0.5	134	2	US-08-777-019-78	Sequence 78, Appl
839	6	0.5	134	2	US-08-606-176A-1	Sequence 1, Appli
840	6	0.5	134	2	US-08-777-143-76	Sequence 76, Appl
841	6	0.5	134	2	US-08-777-143-77	Sequence 77, Appl
842	6	0.5	134	2	US-08-777-143-78	Sequence 78, Appl
843	6	0.5	134	2	US-08-710-219A-1	Sequence 1, Appli
844	6	0.5	134	2	US-08-452-242-4	Sequence 4, Appli
845	6	0.5	134	2	US-08-452-242-6	Sequence 6, Appli
846	6	0.5	134	3	US-08-453-176A-4	Sequence 4, Appli
847	6	0.5	134	3	US-08-453-176A-6	Sequence 6, Appli
848	6	0.5	134	3	US-08-775-414-76	Sequence 76, Appl
849	6	0.5	134	3	US-08-775-414-77	Sequence 77, Appl
850	6	0.5	134	3	US-08-775-414-78	Sequence 78, Appl
851	6	0.5	134	3	US-08-451-374-4	Sequence 4, Appli
852	6	0.5	134	3	US-08-451-374-6	Sequence 6, Appli
853	6	0.5	134	3	US-08-535-681-2	Sequence 2, Appli
854	6	0.5	134	3	US-08-935-268A-4	Sequence 4, Appli
855	6	0.5	134	3	US-08-935-268A-6	Sequence 6, Appli
856	6	0.5	134	3	US-08-931-858E-76	Sequence 76, Appl
857	6	0.5	134	3	US-08-931-858E-77	Sequence 77, Appl
858	6	0.5	134	3	US-08-931-858E-78	Sequence 78, Appl
859	6	0.5	134	3	US-08-981-739-76	Sequence 76, Appl
860	6	0.5	134	3	US-08-981-739-77	Sequence 77, Appl
861	6	0.5	134	3	US-08-981-739-78	Sequence 78, Appl
862	6	0.5	134	3	US-09-220-528-13	Sequence 13, Appl
863	6	0.5	134	4	US-08-452-229-4	Sequence 4, Appli
864	6	0.5	134	4	US-08-452-229-6	Sequence 6, Appli
865	6	0.5	134	4	US-09-128-026-76	Sequence 76, Appl
866	6	0.5	134	4	US-09-128-026-77	Sequence 77, Appl
867	6	0.5	134	4	US-09-128-026-78	Sequence 78, Appl
868	6	0.5	134	4	US-09-732-210-1566	Sequence 1566, Ap
869	6	0.5	135	3	US-08-535-681-47	Sequence 47, Appl
870	6	0.5	135	3	US-08-905-223-464	Sequence 464, App
871	6	0.5	136	4	US-09-107-532A-6435	Sequence 6435, Ap
872	6	0.5	137	4	US-09-252-991A-27555	Sequence 27555, A
873	6	0.5	138	4	US-09-180-422B-12	Sequence 12, Appl
874	6	0.5	138	4	US-09-107-532A-4755	Sequence 4755, Ap
875	6	0.5	138	6	5218099-1	Patent No. 5218099
876	6	0.5	139	3	US-09-367-953B-116	Sequence 116, App
877	6	0.5	139	4	US-09-252-991A-21299	Sequence 21299, A
878	6	0.5	139	4	US-09-328-352-7601	Sequence 7601, Ap
879	6	0.5	140	4	US-09-732-210-1255	Sequence 1255, Ap
880	6	0.5	143	4	US-09-134-001C-4711	Sequence 4711, Ap
881	6	0.5	144	4	US-09-134-001C-3102	Sequence 3102, Ap
882	6	0.5	144	4	US-09-252-991A-26540	Sequence 26540, A
883	6	0.5	145	4	US-09-252-991A-22630	Sequence 22630, A
884	6	0.5	147	4	US-09-134-001C-5392	Sequence 5392, Ap
885	6	0.5	149	4	US-09-198-452A-792	Sequence 792, App
886	6	0.5	151	4	US-09-328-352-6429	Sequence 6429, Ap
887	6	0.5	152	4	US-09-732-210-53	Sequence 53, Appl
888	6	0.5	154	4	US-09-252-991A-17864	Sequence 17864, A
889	6	0.5	154	4	US-09-198-452A-239	Sequence 239, App
890	6	0.5	155	2	US-08-844-057-4	Sequence 4, Appli
891	6	0.5	155	4	US-09-006-730-4	Sequence 4, Appli
892	6	0.5	155	4	US-09-107-532A-5729	Sequence 5729, Ap
893	6	0.5	157	2	US-08-630-822A-60	Sequence 60, Appl
894	6	0.5	157	2	US-09-005-069-60	Sequence 60, Appl
895	6	0.5	157	4	US-09-171-156A-18	Sequence 18, Appl
896	6	0.5	157	4	US-09-004-730A-18	Sequence 18, Appl
897	6	0.5	157	4	US-09-252-991A-22957	Sequence 22957, A
898	6	0.5	157	4	US-08-981-799A-18	Sequence 18, Appl
899	6	0.5	158	4	US-09-433-248A-4	Sequence 4, Appli
900	6	0.5	159	4	US-09-252-991A-16672	Sequence 16672, A
901	6	0.5	159	4	US-09-252-991A-20158	Sequence 20158, A
902	6	0.5	159	4	US-09-252-991A-21183	Sequence 21183, A
903	6	0.5	160	4	US-09-328-352-6929	Sequence 6929, Ap
904	6	0.5	160	4	US-09-732-210-1604	Sequence 1604, Ap
905	6	0.5	161	4	US-09-252-991A-30294	Sequence 30294, A
906	6	0.5	162	4	US-09-198-452A-948	Sequence 948, App
907	6	0.5	164	4	US-09-134-001C-5325	Sequence 5325, Ap
908	6	0.5	164	4	US-09-252-991A-20166	Sequence 20166, A
909	6	0.5	164	4	US-09-252-991A-32517	Sequence 32517, A
910	6	0.5	167	2	US-08-747-887-23	Sequence 23, Appl
911	6	0.5	167	4	US-09-328-352-6884	Sequence 6884, Ap
912	6	0.5	169	1	US-08-414-926A-12	Sequence 12, Appl
913	6	0.5	169	2	US-08-926-922-12	Sequence 12, Appl
914	6	0.5	169	3	US-09-253-682-12	Sequence 12, Appl
915	6	0.5	169	3	US-09-527-657-12	Sequence 12, Appl
916	6	0.5	169	4	US-09-134-001C-4590	Sequence 4590, Ap
917	6	0.5	169	4	US-09-544-716-17	Sequence 17, Appl
918	6	0.5	169	4	US-09-557-921-18	Sequence 18, Appl
919	6	0.5	171	1	US-08-313-075A-32	Sequence 32, Appl
920	6	0.5	171	3	US-09-023-082A-116	Sequence 116, App
921	6	0.5	172	4	US-09-252-991A-28151	Sequence 28151, A
922	6	0.5	174	3	US-09-023-082A-43	Sequence 43, Appl
923	6	0.5	174	3	US-09-023-082A-55	Sequence 55, Appl
924	6	0.5	174	3	US-09-023-082A-66	Sequence 66, Appl
925	6	0.5	174	3	US-08-941-155-2	Sequence 2, Appli
926	6	0.5	174	4	US-08-936-165A-451	Sequence 451, App
927	6	0.5	174	4	US-09-872-047-2	Sequence 2, Appli
928	6	0.5	176	4	US-09-252-991A-21613	Sequence 21613, A
929	6	0.5	178	2	US-08-680-326-4	Sequence 4, Appli
930	6	0.5	178	4	US-09-328-352-6897	Sequence 6897, Ap
931	6	0.5	179	4	US-09-328-352-7612	Sequence 7612, Ap
932	6	0.5	181	1	US-08-185-432-6	Sequence 6, Appli
933	6	0.5	182	4	US-09-134-001C-5022	Sequence 5022, Ap
934	6	0.5	182	4	US-09-328-352-6838	Sequence 6838, Ap
935	6	0.5	184	3	US-09-023-082A-30	Sequence 30, Appl
936	6	0.5	184	3	US-08-741-411-1	Sequence 1, Appli
937	6	0.5	185	1	US-07-741-940-6	Sequence 6, Appli
938	6	0.5	185	1	US-08-289-548A-6	Sequence 6, Appli
939	6	0.5	185	1	US-08-452-654-6	Sequence 6, Appli
940	6	0.5	185	1	US-08-452-655B-6	Sequence 6, Appli
941	6	0.5	185	2	US-08-865-336-4	Sequence 4, Appli
942	6	0.5	185	3	US-08-450-582-6	Sequence 6, Appli
943	6	0.5	185	4	US-08-449-731-6	Sequence 6, Appli
944	6	0.5	185	4	US-09-996-243-211	Sequence 211, App
945	6	0.5	186	4	US-09-252-991A-30615	Sequence 30615, A
946	6	0.5	186	4	US-09-107-532A-7101	Sequence 7101, Ap
947	6	0.5	189	4	US-09-107-532A-4526	Sequence 4526, Ap
948	6	0.5	191	4	US-09-252-991A-29829	Sequence 29829, A
949	6	0.5	193	4	US-08-671-548C-40	Sequence 40, Appl
950	6	0.5	194	4	US-09-134-001C-3563	Sequence 3563, Ap
951	6	0.5	194	4	US-09-149-476-735	Sequence 735, App
952	6	0.5	194	4	US-09-252-991A-21846	Sequence 21846, A
953	6	0.5	195	4	US-09-487-792-11	Sequence 11, Appl
954	6	0.5	195	4	US-09-908-594-11	Sequence 11, Appl
955	6	0.5	196	4	US-09-252-991A-24314	Sequence 24314, A
956	6	0.5	197	4	US-09-328-352-5000	Sequence 5000, Ap
957	6	0.5	199	3	US-09-103-359-1	Sequence 1, Appli
958	6	0.5	200	3	US-09-113-750A-40	Sequence 40, Appl
959	6	0.5	200	4	US-09-562-737-101	Sequence 101, App
960	6	0.5	200	4	US-09-562-737-102	Sequence 102, App
961	6	0.5	200	4	US-09-562-737-109	Sequence 109, App
962	6	0.5	200	4	US-09-562-737-110	Sequence 110, App
963	6	0.5	200	4	US-09-252-991A-27855	Sequence 27855, A
964	6	0.5	204	2	US-08-760-075A-16	Sequence 16, Appl
965	6	0.5	204	3	US-09-338-546-16	Sequence 16, Appl
966	6	0.5	204	4	US-09-659-084-16	Sequence 16, Appl
967	6	0.5	205	4	US-09-252-991A-18239	Sequence 18239, A
968	6	0.5	205	4	US-09-198-452A-829	Sequence 829, App
969	6	0.5	206	4	US-09-134-001C-5047	Sequence 5047, Ap
970	6	0.5	207	4	US-09-252-991A-29505	Sequence 29505, A
971	6	0.5	207	4	US-09-434-354-40	Sequence 40, Appl
972	6	0.5	208	1	US-08-462-169B-20	Sequence 20, Appl
973	6	0.5	208	2	US-08-951-822-30	Sequence 30, Appl
974	6	0.5	208	3	US-09-103-079-20	Sequence 20, Appl
975	6	0.5	208	3	US-09-023-082A-2	Sequence 2, Appli
976	6	0.5	208	3	US-09-023-082A-20	Sequence 20, Appl

```
  977     6    0.5    208  3  US-09-023-082A-24         Sequence 24, Appl
  978     6    0.5    208  3  US-09-023-082A-39         Sequence 39, Appl
  979     6    0.5    208  3  US-09-218-444-2           Sequence 2, Appli
  980     6    0.5    208  4  US-09-368-951-30          Sequence 30, Appl
  981     6    0.5    208  4  US-09-425-021-20          Sequence 20, Appl
  982     6    0.5    208  4  US-09-229-947-30          Sequence 30, Appl
  983     6    0.5    208  4  US-09-252-991A-21959      Sequence 21959, A
  984     6    0.5    209  4  US-08-914-999-14          Sequence 14, Appl
  985     6    0.5    209  4  US-09-198-452A-286        Sequence 286, App
  986     6    0.5    210  1  US-07-741-940-4           Sequence 4, Appli
  987     6    0.5    210  1  US-08-289-548A-4          Sequence 4, Appli
  988     6    0.5    210  1  US-08-452-654-4           Sequence 4, Appli
  989     6    0.5    210  1  US-08-452-655B-4          Sequence 4, Appli
  990     6    0.5    210  3  US-08-450-582-4           Sequence 4, Appli
  991     6    0.5    210  4  US-09-134-001C-3671       Sequence 3671, Ap
  992     6    0.5    210  4  US-08-449-731-4           Sequence 4, Appli
  993     6    0.5    217  4  US-08-679-493A-196        Sequence 196, App
  994     6    0.5    218  4  US-09-252-991A-22136      Sequence 22136, A
  995     6    0.5    220  1  US-07-820-154A-4          Sequence 4, Appli
  996     6    0.5    220  2  US-08-097-554A-4          Sequence 4, Appli
  997     6    0.5    220  3  US-08-453-176A-26         Sequence 26, Appl
  998     6    0.5    220  3  US-08-480-640A-4          Sequence 4, Appli
  999     6    0.5    220  3  US-08-451-374-26          Sequence 26, Appl
 1000     6    0.5    220  3  US-08-295-802-4           Sequence 4, Appli

                              ALIGNMENTS


RESULT 1
US-09-384-162-6
; Sequence 6, Application US/09384162
; Patent No. 6376747
; GENERAL INFORMATION:
;  APPLICANT: Xing, Ti
;  APPLICANT: Malik, Kamal
;  APPLICANT: Martin-Heller, Teresa
;  APPLICANT: Miki L., Brian
;  TITLE OF INVENTION: No. 6376747el Plant-Derived Map Kinase Kinase
;  FILE REFERENCE: 08-884280US
;  CURRENT APPLICATION NUMBER: US/09/384,162
;  CURRENT FILING DATE: 1999-08-27
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 286
;   TYPE: PRT
;   ORGANISM: Dictyostelium discoideum
US-09-384-162-6

  Query Match             1.2%;  Score 16;  DB 4;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        206 NNNNNNNNNNNNNNNN 221

RESULT 2
US-09-718-841-4
; Sequence 4, Application US/09718841
; Patent No. 6333184
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6333184el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1051
;  CURRENT APPLICATION NUMBER: US/09/718,841
;  CURRENT FILING DATE:  2000-11-22
;  NUMBER OF SEQ ID NCS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 357
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (16)...(37)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-718-841-4

  Query Match             1.2%;  Score 16;  DB 4;  Length 357;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTMLG 107
              ||||||||||||||||
Db         98 FAYGQTGSGKTYTMLG 113


RESULT 3
US-09-718-810-4
; Sequence 4, Application US/09718810
; Patent No. 6420162
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6420162el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1051
;  CURRENT APPLICATION NUMBER: US/09/718,810
;  CURRENT FILING DATE:  2000-11-22
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 357
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (16)...(37)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-718-810-4

  Query Match             1.2%;  Score 16;  DB 4;  Length 357;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTMLG 107
              ||||||||||||||||
Db         98 FAYGQTGSGKTYTMLG 113


RESULT 4
US-09-118-319-5
; Sequence 5, Application US/09118319
; Patent No. 6114158
; GENERAL INFORMATION:
;  APPLICANT: Li, Xin-Liang
;  APPLICANT: Chen, Huizhong
;  APPLICANT: Ljungdahl, Lars G.
;  TITLE OF INVENTION: Orpinomyces Cellulase CelF Protein and Coding Sequences
;  FILE REFERENCE: 33-98sequence listing
;  CURRENT APPLICATION NUMBER: US/09/118,319
;  CURRENT FILING DATE: 1998-07-17
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 428
;   TYPE: PRT
;   ORGANISM: Neocallimastix patriciarum
US-09-118-319-5

  Query Match             1.2%;  Score 16;  DB 3;  Length 428;
```

```
  Best Local Similarity   100.0%;  Pred. No. 3.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db         68 NNNNNNNNNNNNNNNN 83


RESULT 5
US-09-118-319-2
; Sequence 2, Application US/09118319
; Patent No. 6114158
; GENERAL INFORMATION:
;  APPLICANT: Li, Xin-Liang
;  APPLICANT: Chen, Huizhong
;  APPLICANT: Ljungdahl, Lars G.
;  TITLE OF INVENTION: Orpinomyces Cellulase CelF Protein and Coding Sequences
;  FILE REFERENCE: 33-98sequence listing
;  CURRENT APPLICATION NUMBER: US/09/118,319
;  CURRENT FILING DATE: 1998-07-17
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: Orpinomyces sp. PC-2
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:oligonucleotide
US-09-118-319-2

  Query Match              1.2%;  Score 16;  DB 3;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db         78 NNNNNNNNNNNNNNNN 93


RESULT 6
US-09-718-841-2
; Sequence 2, Application US/09718841
; Patent No. 6333184
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6333184el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1051
;  CURRENT APPLICATION NUMBER: US/09/718,841
;  CURRENT FILING DATE:  2000-11-22
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 563
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (16)...(37)
;   OTHER INFORMATION: Xaa = any amino acid
;   NAME/KEY: VARIANT
;   LOCATION: (391)...(412)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-718-841-2

  Query Match              1.2%;  Score 16;  DB 4;  Length 563;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTMLG 107
              ||||||||||||||||
Db         98 FAYGQTGSGKTYTMLG 113
```

```
RESULT 7
US-09-718-810-2
; Sequence 2, Application US/09718810
; Patent No. 6420162
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6420162el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1051
;  CURRENT APPLICATION NUMBER: US/09/718,810
;  CURRENT FILING DATE:  2000-11-22
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 563
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (16)...(37)
;   OTHER INFORMATION: Xaa = any amino acid
;   NAME/KEY: VARIANT
;   LOCATION: (391)...(412)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-718-810-2

  Query Match              1.2%;  Score 16;  DB 4;  Length 563;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTMLG 107
              ||||||||||||||||
Db         98 FAYGQTGSGKTYTMLG 113


RESULT 8
US-08-914-999-8
; Sequence 8, Application US/08914999
; Patent No. 6346406
;  GENERAL INFORMATION:
;    APPLICANT:  Ryazanov, Alexey G.
;    APPLICANT:  Hait, William N.
;    APPLICANT:  Pavur, Karen S.
;    TITLE OF INVENTION:  ELONGATION FACTOR-2 KINASE (EF-2 KINASE)
;    TITLE OF INVENTION:  AND METHODS OF USE THEREFOR
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David A. Jackson, Esq.
;      STREET:  411 Hackensack Ave, Continental Plaza, 4th
;      STREET:  Floor
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/914,999
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  601-1-078
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-487-5800
;      TELEFAX:  201-343-1684
```

```
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  732 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Dictyostelium discoideum
US-08-914-999-8

  Query Match             1.2%;  Score 16;  DB 4;  Length 732;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-07;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        355 NNNNNNNNNNNNNNNN 370

RESULT 9
US-08-340-011-5
; Sequence 5, Application US/08340011
; Patent No. 5776755
;  GENERAL INFORMATION:
;    APPLICANT:  Alitalo, et al.
;    TITLE OF INVENTION:  FLT4, A NOVEL RECEPTOR TYROSIN KINASE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/340,011
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959,951
;      FILING DATE:  09-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Thomas C.
;      REGISTRATION NUMBER:  36,989
;      REFERENCE/DOCKET NUMBER:  32267
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1311 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-340-011-5

  Query Match             1.2%;  Score 16;  DB 1;  Length 1311;
  Best Local Similarity   100.0%;  Pred. No. 1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        613 NNNNNNNNNNNNNNNN 628

RESULT 10
US-08-901-710-5
; Sequence 5, Application US/08901710
; Patent No. 6107046
;  GENERAL INFORMATION:
;    APPLICANT:  Alitalo, Kari
;    APPLICANT:  Aprelikova, Olga
;    APPLICANT:  Pajusola, Katri
;    APPLICANT:  Armstrong, Elina
;    APPLICANT:  Korhonen, Jaana
;    APPLICANT:  Kaipainen, Arja
;    APPLICANT:  Matikainen, Marja-Terttu
;    TITLE OF INVENTION:  FLT4, A RECEPTOR TYROSINE KINASE, AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/901,710
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/340,011
;      FILING DATE:  14-NOV-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/257,754
;      FILING DATE:  09-JUL-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959,951
;      FILING DATE:  09-OCT-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,153
;      REFERENCE/DOCKET NUMBER:  28113/33824
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1311 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-901-710-5

  Query Match             1.2%;  Score 16;  DB 3;  Length 1311;
  Best Local Similarity   100.0%;  Pred. No. 1e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        613 NNNNNNNNNNNNNNNN 628

RESULT 11
US-09-457-040B-27
; Sequence 27, Application US/09457040B
; Patent No. 6387641
; GENERAL INFORMATION:
;  APPLICANT: Vertex Pharmaceuticals Incorporated
```

```
;  APPLICANT:  Bellon, Steve
;  TITLE OF INVENTION: Crystallized P38 Complexes
;  FILE REFERENCE: VPI/98-14
;  CURRENT APPLICATION NUMBER: US/09/457,040B
;  CURRENT FILING DATE:  1999-12-08
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
;   LENGTH: 1584
;   TYPE: PRT
;   ORGANISM: DICDI - Dictyostelium Discoideum
US-09-457-040B-27

  Query Match             1.2%;  Score 16;  DB 4;  Length 1584;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        449 NNNNNNNNNNNNNNNN 464

RESULT 12
US-09-417-485D-6
; Sequence 6, Application US/09417485D
; Patent No. 6541202
; GENERAL INFORMATION:
;  APPLICANT: Long, David M.
;  APPLICANT:  Metz, Anneke M.
;  APPLICANT:  Love, Ruschelle A.
;  TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
;  FILE REFERENCE: 47714-5009-US
;  CURRENT APPLICATION NUMBER: US/09/417,485D
;  CURRENT FILING DATE:  2002-06-14
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 2184
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (330)..(335)
;   OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
;   OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
US-09-417-485D-6

  Query Match             1.2%;  Score 16;  DB 4;  Length 2184;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-06;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNNN 757
              ||||||||||||||||
Db        239 NNNNNNNNNNNNNNNN 254

RESULT 13
US-09-722-129-4
; Sequence 4, Application US/09722129
; Patent No. 6582958
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6582958el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1054
;  CURRENT APPLICATION NUMBER: US/09/722,129
;  CURRENT FILING DATE:  2000-11-24
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 303
;   TYPE: PRT
;   ORGANISM: Human
US-09-722-129-4

  Query Match             1.2%;  Score 15;  DB 4;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         91 CFAYGQTGSGKTYTM 105
              |||||||||||||||
Db         90 CFAYGQTGSGKTYTM 104

RESULT 14
US-09-722-129-2
; Sequence 2, Application US/09722129
; Patent No. 6582958
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6582958el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1054
;  CURRENT APPLICATION NUMBER: US/09/722,129
;  CURRENT FILING DATE:  2000-11-24
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 492
;   TYPE: PRT
;   ORGANISM: Human
US-09-722-129-2

  Query Match             1.2%;  Score 15;  DB 4;  Length 492;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         91 CFAYGQTGSGKTYTM 105
              |||||||||||||||
Db        279 CFAYGQTGSGKTYTM 293

RESULT 15
US-09-417-485D-8
; Sequence 8, Application US/09417485D
; Patent No. 6541202
; GENERAL INFORMATION:
;  APPLICANT: Long, David M.
;  APPLICANT:  Metz, Anneke M.
;  APPLICANT:  Love, Ruschelle A.
;  TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
;  FILE REFERENCE: 47714-5009-US
;  CURRENT APPLICATION NUMBER: US/09/417,485D
;  CURRENT FILING DATE:  2002-06-14
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 794
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-09-417-485D-8

  Query Match             1.2%;  Score 15;  DB 4;  Length 794;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        744 NNNNNNNNNNNNNNI 758
              |||||||||||||||
Db        130 NNNNNNNNNNNNNNI 144

RESULT 16
US-08-861-464-6
; Sequence 6, Application US/08861464
```

```
; Patent No. 5874210
;  GENERAL INFORMATION:
;    APPLICANT:  Guarente, Leonard P.
;    APPLICANT:  Austriaco Jr., Nicanor
;    APPLICANT:  Kennedy, Brian
;    TITLE OF INVENTION:  Genes Determining Cellular Senescence
;    TITLE OF INVENTION:  in Yeast
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/861,464
;      FILING DATE:  22-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/396,001
;      FILING DATE:  28-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/09351
;      FILING DATE:  15-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/107,408
;      FILING DATE:  16-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-6408A2Z
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  781-861-6240
;      TELEFAX:  781-861-9540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  888 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-861-464-6

  Query Match             1.2%;  Score 15;  DB 2;  Length 888;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNN 756
              |||||||||||||||
Db        521 NNNNNNNNNNNNNNN 535

RESULT 17
US-08-396-001-6
; Sequence 6, Application US/08396001
; Patent No. 5919618
;  GENERAL INFORMATION:
;    APPLICANT:  Guarente, Leonard P.
;    APPLICANT:  Austriaco Jr., Nicanor
;    APPLICANT:  Claus, James
;    APPLICANT:  Cole, Francesca
;    APPLICANT:  Kennedy, Brian
;    TITLE OF INVENTION:  Genes Determining Cellular Senesence in
;    TITLE OF INVENTION:  Yeast
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/396,001
;      FILING DATE:  28-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-6408A2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  888 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-396-001-6

  Query Match             1.2%;  Score 15;  DB 2;  Length 888;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNN 756
              |||||||||||||||
Db        521 NNNNNNNNNNNNNNN 535

RESULT 18
US-09-323-433A-6
; Sequence 6, Application US/09323433A
; Patent No. 6218512
; GENERAL INFORMATION:
;  APPLICANT:  Guarente, Leonard P.
;  APPLICANT:  Austriaco Jr., Nicanor
;  APPLICANT:  Claus, James J.
;  APPLICANT:  Cole, Francesca
;  APPLICANT:  Kennedy, Brian
;  TITLE OF INVENTION:  GENES DETERMINING CELLULAR SENESCENCE IN
;  TITLE OF INVENTION:  YEAST
;  FILE REFERENCE:  0050.1491-003
;  CURRENT APPLICATION NUMBER: US/09/323,433A
;  CURRENT FILING DATE:  1999-06-01
;  PRIOR APPLICATION NUMBER:  US 08/396,001
;  PRIOR FILING DATE:  1995-02-28
;  PRIOR APPLICATION NUMBER:  PCT/US94/09351
;  PRIOR FILING DATE:  1994-08-15
;  PRIOR APPLICATION NUMBER:  US 08/107,408
;  PRIOR FILING DATE:  1993-08-16
;  NUMBER OF SEQ ID NOS:  48
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 888
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-323-433A-6

  Query Match             1.2%;  Score 15;  DB 3;  Length 888;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        742 NNNNNNNNNNNNNNN 756
              |||||||||||||||
Db        521 NNNNNNNNNNNNNNN 535
```

```
RESULT 19
US-09-098-901-11
; Sequence 11, Application US/09098901B
; Patent No. 6218144
; GENERAL INFORMATION:
;  APPLICANT: Scott, Matthew
;  APPLICANT: Sisson, John C.
;  TITLE OF INVENTION: Costal2 Genes and their Uses
;  FILE REFERENCE: SUN-65P
;  CURRENT APPLICATION NUMBER: US/09/098,901B
;  CURRENT FILING DATE: 1998-06-17
;  EARLIER APPLICATION NUMBER: 60/051,347
;  EARLIER FILING DATE: 1997-06-30
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
;   LENGTH: 23
;   TYPE: PRT
;   ORGANISM: CONSENSUS
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(23)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-098-901-11

  Query Match             1.1%;  Score 14;  DB 3;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db          8 FAYGQTGSGKTYTM 21

RESULT 20
US-09-641-806-4
; Sequence 4, Application US/09641806
; Patent No. 6395527
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6395527el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1034
;  CURRENT APPLICATION NUMBER: US/09/641,806
;  CURRENT FILING DATE:  2000-08-17
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Human
US-09-641-806-4

  Query Match             1.1%;  Score 14;  DB 4;  Length 337;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db         80 FAYGQTGSGKTYTM 93

RESULT 21
US-09-723-129-4
; Sequence 4, Application US/09723129
; Patent No. 6551787
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6551787el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1034
;  CURRENT APPLICATION NUMBER: US/09/723,129
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 09/641,806
;  PRIOR FILING DATE: 2000-08-17
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Human
US-09-723-129-4

  Query Match             1.1%;  Score 14;  DB 4;  Length 337;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db         80 FAYGQTGSGKTYTM 93

RESULT 22
US-09-722-862-4
; Sequence 4, Application US/09722862
; Patent No. 6562610
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6562610el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1034
;  CURRENT APPLICATION NUMBER: US/09/722,862
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 09/641,806
;  PRIOR FILING DATE: 2000-08-17
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Human
US-09-722-862-4

  Query Match             1.1%;  Score 14;  DB 4;  Length 337;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db         80 FAYGQTGSGKTYTM 93

RESULT 23
US-09-641-806-2
; Sequence 2, Application US/09641806
; Patent No. 6395527
; GENERAL INFORMATION:
;  APPLICANT:  Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION:  No. 6395527el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE:  1034
;  CURRENT APPLICATION NUMBER: US/09/641,806
;  CURRENT FILING DATE:  2000-08-17
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 342
;   TYPE: PRT
;   ORGANISM: Human
US-09-641-806-2
```

```
  Query Match             1.1%;  Score 14;  DB 4;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db         81 FAYGQTGSGKTYTM 94

RESULT 24
US-09-723-129-2
; Sequence 2, Application US/09723129
; Patent No. 6551787
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6551787el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1034
;  CURRENT APPLICATION NUMBER: US/09/723,129
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 09/641,806
;  PRIOR FILING DATE: 2000-08-17
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 342
;   TYPE: PRT
;   ORGANISM: Human
US-09-723-129-2

  Query Match             1.1%;  Score 14;  DB 4;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db         81 FAYGQTGSGKTYTM 94

RESULT 25
US-09-722-862-2
; Sequence 2, Application US/09722862
; Patent No. 6562610
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6562610el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1034
;  CURRENT APPLICATION NUMBER: US/09/722,862
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 09/641,806
;  PRIOR FILING DATE: 2000-08-17
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 342
;   TYPE: PRT
;   ORGANISM: Human
US-09-722-862-2

  Query Match             1.1%;  Score 14;  DB 4;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db         81 FAYGQTGSGKTYTM 94

RESULT 26
US-09-291-170A-3
; Sequence 3, Application US/09291170A
; Patent No. 6410687
; GENERAL INFORMATION:
;  APPLICANT:  Vale, Ronald D.
;  APPLICANT:  Hartman, James J.
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION:  Assays for the Detection of Microtubule
;  TITLE OF INVENTION:  Depolymerization Inhibitors
;  FILE REFERENCE:  18557B-000510US
;  CURRENT APPLICATION NUMBER: US/09/291,170A
;  CURRENT FILING DATE:  1999-04-13
;  PRIOR APPLICATION NUMBER:  US 60/081,734
;  PRIOR FILING DATE:  1998-04-14
;  NUMBER OF SEQ ID NOS:  16
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 730
;   TYPE: PRT
;   ORGANISM: Xenopus laevis
;   FEATURE:
;   OTHER INFORMATION: Xenopus kinesin central motor 1 (XKCM1)
US-09-291-170A-3

  Query Match             1.1%;  Score 14;  DB 4;  Length 730;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 INRSLLALKECIRA 276
              ||||||||||||||
Db        520 INRSLLALKECIRA 533

RESULT 27
US-09-724-884-3
; Sequence 3, Application US/09724884
; Patent No. 6429304
; GENERAL INFORMATION:
;  APPLICANT: Vale, Ronald D.
;  APPLICANT:  Hartman, James J.
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: Assays for the Detection of Microtubule
;  TITLE OF INVENTION:  Depolymerization Inhibitors
;  FILE REFERENCE: 18557B-000510US
;  CURRENT APPLICATION NUMBER: US/09/724,884
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER: 09/291,170
;  PRIOR FILING DATE: 1999-04-13
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 730
;   TYPE: PRT
;   ORGANISM: Xenopus laevis
;   FEATURE:
;   OTHER INFORMATION: Xenopus kinesin central motor 1 (XKCM1)
US-09-724-884-3

  Query Match             1.1%;  Score 14;  DB 4;  Length 730;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        263 INRSLLALKECIRA 276
              ||||||||||||||
Db        520 INRSLLALKECIRA 533

RESULT 28
US-08-006-676B-1
; Sequence 1, Application US/08006676B
; Patent No. 5411865
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven
```

```
;    TITLE OF INVENTION:  Diagnosis of Leishmaniasis
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jeffrey B. Oster
;      STREET:  8339 SE 57th Street
;      CITY:  Mercer Island
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98040-4906
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatable
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORD for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/006,676B
;      FILING DATE:  15-JAN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oster, Jeffrey B.
;      REGISTRATION NUMBER:  32,585
;      REFERENCE/DOCKET NUMBER:  REED-4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 232 7845
;      TELEFAX:  (206) 236 0205
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  955 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-006-676B-1

  Query Match             1.1%;  Score 14;  DB 1;  Length 955;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        119 FAYGQTGSGKTYTM 132


RESULT 29
US-08-282-845-2
; Sequence 2, Application US/08282845
; Patent No. 5719263
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  A 230Kd Antigen Present in Leishmania
;    TITLE OF INVENTION:  Species
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple Macintosh Operating System 7.1
;      SOFTWARE:  Microsoft Word for Macintosh 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/282,845
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/006,676
;      FILING DATE:  JANUARY 15, 1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia Anne
;      REGISTRATION NUMBER:  34,693
;      REFERENCE/DOCKET NUMBER:  5004-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206)587-0430
;      TELEFAX:  (206)233-0644
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  955 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-282-845-2

  Query Match             1.1%;  Score 14;  DB 1;  Length 955;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        119 FAYGQTGSGKTYTM 132


RESULT 30
US-08-428-414A-3
; Sequence 3, Application US/08428414A
; Patent No. 5912166
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF
;    TITLE OF INVENTION:  LEISHMANIASIS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/428,414A
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kadlecek, Ann T.
;      REGISTRATION NUMBER:  39,244
;      REFERENCE/DOCKET NUMBER:  210121.407
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANDBERRY
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  955 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-428-414A-3

  Query Match             1.1%;  Score 14;  DB 2;  Length 955;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        119 FAYGQTGSGKTYTM 132


RESULT 31
```

```
PCT-US94-00324-1
; Sequence 1, Application PC/TUS9400324
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven
;    TITLE OF INVENTION:  Diagnosis of Leishmaniasis
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple System 7.1
;      SOFTWARE:  Microsoft Word, version 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/00324
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/006,676
;      FILING DATE:  15-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia Anne
;      REGISTRATION NUMBER:  34,693
;      REFERENCE/DOCKET NUMBER:  5004-WO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 587-0430
;      TELEFAX:  (206) 233-0644
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  955 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US94-00324-1

  Query Match             1.1%;  Score 14;  DB 5;  Length 955;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-05;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYTM 105
              ||||||||||||||
Db        119 FAYGQTGSGKTYTM 132

RESULT 32
US-09-724-517-4
; Sequence 4, Application US/09724517
; Patent No. 6379941
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6379941el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1031
;  CURRENT APPLICATION NUMBER: US/09/724,517
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: US/09/641,807
;  PRIOR FILING DATE: 2000-08-17
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: Human
US-09-724-517-4

  Query Match             1.0%;  Score 13;  DB 4;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 0.00027;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYT 104
              |||||||||||||
Db         80 FAYGQTGSGKTYT 92

RESULT 33
US-09-641-807A-4
; Sequence 4, Application US/09641807A
; Patent No. 6440731
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6440731el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1031
;  CURRENT APPLICATION NUMBER: US/09/641,807A
;  CURRENT FILING DATE:  2000-08-17
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: Human
US-09-641-807A-4

  Query Match             1.0%;  Score 13;  DB 4;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 0.00027;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYT 104
              |||||||||||||
Db         80 FAYGQTGSGKTYT 92

RESULT 34
US-09-723-096-4
; Sequence 4, Application US/09723096
; Patent No. 6448026
; GENERAL INFORMATION:
;  APPLICANT: Beraud, Christophe
;  APPLICANT:  Freedman, Richard
;  TITLE OF INVENTION: No. 6448026el motor proteins and methods for
;  TITLE OF INVENTION:  their use
;  FILE REFERENCE: 1031
;  CURRENT APPLICATION NUMBER: US/09/723,096
;  CURRENT FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: US/09/641,807
;  PRIOR FILING DATE: 2000-08-17
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 341
;   TYPE: PRT
;   ORGANISM: Human
US-09-723-096-4

  Query Match             1.0%;  Score 13;  DB 4;  Length 341;
  Best Local Similarity   100.0%;  Pred. No. 0.00027;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYT 104
              |||||||||||||
Db         80 FAYGQTGSGKTYT 92

RESULT 35
US-09-177-165A-31
; Sequence 31, Application US/09177165A
; Patent No. 6426205
; GENERAL INFORMATION:
;  APPLICANT: Tyers, Mike
```

```
; APPLICANT: Willems, Andrew
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING UBIQUITIN
; TITLE OF INVENTION: DEPENDENT PROTEOLYSIS
; FILE REFERENCE: 11757.10USU1
; CURRENT APPLICATION NUMBER: US/09/177,165A
; CURRENT FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/092,443
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 60/063,254
; PRIOR FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
;   LENGTH: 1151
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-177-165A-31

  Query Match             1.0%;  Score 13;  DB 4;  Length 1151;
  Best Local Similarity   100.0%;  Pred. No. 0.0008;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        502 NNNNNNNNNNNNS 514
              |||||||||||||
Db         38 NNNNNNNNNNNNS 50

RESULT 36
US-09-724-517-2
; Sequence 2, Application US/09724517
; Patent No. 6379941
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6379941el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1031
; CURRENT APPLICATION NUMBER: US/09/724,517
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US/09/641,807
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1279
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (409)...(436)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-724-517-2

  Query Match             1.0%;  Score 13;  DB 4;  Length 1279;
  Best Local Similarity   100.0%;  Pred. No. 0.00088;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYT 104
              |||||||||||||
Db         82 FAYGQTGSGKTYT 94

RESULT 37
US-09-641-807A-2
; Sequence 2, Application US/09641807A
; Patent No. 6440731
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6440731el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1031
; CURRENT APPLICATION NUMBER: US/09/641,807A
; CURRENT FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1279
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (409)...(446)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-641-807A-2

  Query Match             1.0%;  Score 13;  DB 4;  Length 1279;
  Best Local Similarity   100.0%;  Pred. No. 0.00088;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYT 104
              |||||||||||||
Db         82 FAYGQTGSGKTYT 94

RESULT 38
US-09-723-096-2
; Sequence 2, Application US/09723096
; Patent No. 6448026
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6448026el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1031
; CURRENT APPLICATION NUMBER: US/09/723,096
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US/09/641,807
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1279
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (409)...(436)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-723-096-2

  Query Match             1.0%;  Score 13;  DB 4;  Length 1279;
  Best Local Similarity   100.0%;  Pred. No. 0.00088;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         92 FAYGQTGSGKTYT 104
              |||||||||||||
Db         82 FAYGQTGSGKTYT 94

RESULT 39
US-09-451-117-2
; Sequence 2, Application US/09451117
; Patent No. 6277973
; GENERAL INFORMATION:
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Fayer, Ronald
; APPLICANT: Trout, James
; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41
; Patent No. 6277973
; TITLE OF INVENTION: kDa Cryptosporidium parvum Oocyst Wall Protein
; FILE REFERENCE: 0046.99
; CURRENT APPLICATION NUMBER: US/09/451,117
; CURRENT FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 2
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Cryptosporidium parvum
US-09-451-117-2

  Query Match             0.9%;  Score 12;  DB 3;  Length 246;
  Best Local Similarity   100.0%;  Pred. No. 0.0019;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNN 752
              ||||||||||||
Db        235 TNNNNNNNNNNN 246


RESULT 40
US-09-888-655-2
; Sequence 2, Application US/09888655
; Patent No. 6521229
; GENERAL INFORMATION:
;  APPLICANT: Jenkins, Mark C.
;  APPLICANT:  Fayer, Ronald
;  APPLICANT:  Trout, James
;  TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41
; Patent No. 6521229
;  TITLE OF INVENTION:  kDa Cryptosporidium parvum Oocyst Wall Protein
;  FILE REFERENCE: 0046.99
;  CURRENT APPLICATION NUMBER: US/09/888,655
;  CURRENT FILING DATE:  2001-06-26
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Cryptosporidium parvum
US-09-888-655-2

  Query Match             0.9%;  Score 12;  DB 4;  Length 246;
  Best Local Similarity   100.0%;  Pred. No. 0.0019;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        741 TNNNNNNNNNNN 752
              ||||||||||||
Db        235 TNNNNNNNNNNN 246


Search completed: October  2, 2003, 16:36:31
Job time : 36 secs
```